Justification of the Nonlinear Schrödinger equation for the evolution of gravity driven 2D surface water waves in a canal of finite depth

Wolf-Patrick Düll[1], Guido Schneider[1] and C. Eugene Wayne[2]

June 13, 2012

Abstract

In 1968 V.E. Zakharov derived the Nonlinear Schrödinger equation for the 2D water wave problem in the absence of surface tension, i.e., for the evolution of gravity driven surface water waves, in order to describe slow temporal and spatial modulations of a spatially and temporarily oscillating wave packet. In this paper we give a rigorous proof that the wave packets in the two-dimensional water wave problem in a canal of finite depth can be accurately approximated by solutions of the Nonlinear Schrödinger equation.

Contents

[1]Institut für Analysis, Dynamik und Modellierung, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, duell@mathematik.uni-stuttgart.de, guido.schneider@mathematik.uni-stuttgart.de

[2]Dept. of Mathematics and Statistics and Center for Biodynamics, Boston University, 111 Cummington Street, Boston, MA 02215, USA, cew@math.bu.edu

1 Introduction

In 1968 V.E. Zakharov [Za68] derived the Nonlinear Schrödinger (NLS) equation

$$\partial_T A = i\nu_1 \partial_X^2 A + i\nu_2 A|A|^2, \tag{1}$$

with $T \in \mathbb{R}$, $X \in \mathbb{R}$, $A(X,T) \in \mathbb{C}$, and coefficients $\nu_j = \nu_j(k_0) \in \mathbb{R}$ from the equations of the 2D water wave problem in case of no surface tension in order to describe slow spatial and temporal modulations of a spatially and temporarily oscillating wave packet $e^{i(k_0 x - \omega_0 t)}$ with a basic spatial wave number $k_0 \neq 0$ and a basic temporal wave number $\omega_0 \neq 0$.

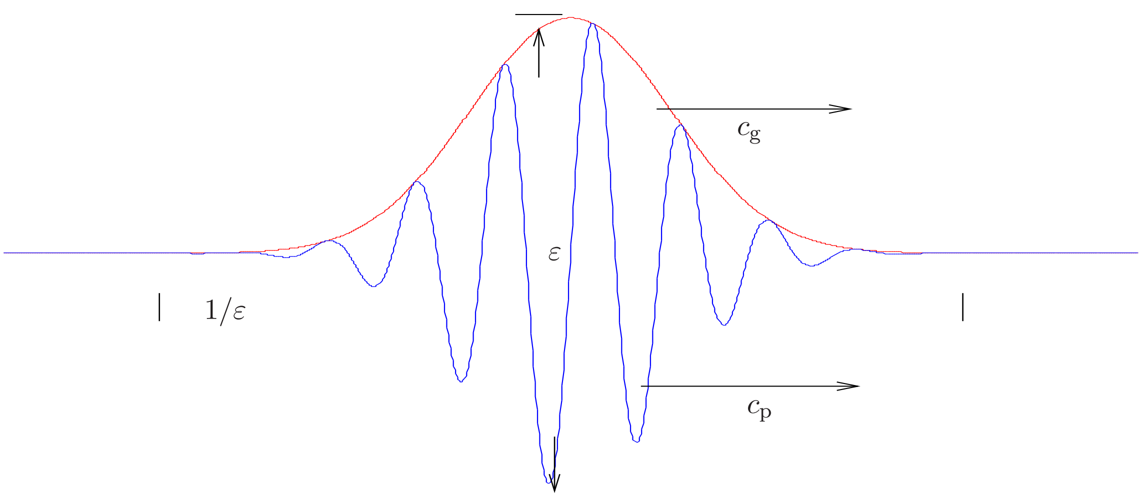

Figure 1: The envelope (advancing with the group velocity c_g) of the oscillating wave packet (advancing with the phase velocity $c_p = \omega_0/k_0$) is described by the amplitude A which solves the NLS equation (1).

The 2D water wave problem without surface tension consists in finding the irrotational flow of an incompressible fluid in an infinitely long canal of finite or infinite depth with a free surface under the influence of gravity. We will consider the case of finite depth. The coordinates are denoted with $x_1 \in \mathbb{R}$ in the horizontal and $x_2 \geq -1$ in the vertical direction. The fluid is contained in the unbounded domain $\Omega(t)$ between the impermeable bottom $\{(x_1,-1)|x_1 \in \mathbb{R}\}$ and the free unknown top surface $\Gamma(t) = \{(x_1, \eta(x_1,t))| \ x_1 \in \mathbb{R}\}$. Under these assumptions it turns out that the problem is completely determined by the evolution of the free surface $\Gamma(t)$.

In detail, the velocity field $u = (u_1, u_2)$ satisfies Euler's equations in $\Omega(t)$. From the assumption of the irrotationality of the flow, i.e., rot $u = 0$, which is preserved by Euler's equations, it follows that the velocity field can be written as a gradient of a potential ϕ :

$\Omega(t) \to I\!R$, i.e., $u = \nabla\phi$, which due to the incompressibility of the fluid, (i.e., div $u = 0$), satisfies the equation

$$\Delta\phi = 0, \qquad \text{in} \quad \Omega(t). \tag{2}$$

The impermeability of the bottom gives the lower boundary condition

$$u_2|_{x_2=-1} = \partial_{x_2}\phi|_{x_2=-1} = 0. \tag{3}$$

On the free surface $\Gamma(t)$ we have the kinematic boundary condition and the balance of forces,

$$\partial_t\eta = \partial_{x_2}\phi - (\partial_{x_1}\eta)\,(\partial_{x_1}\phi), \tag{4}$$

$$\partial_t\phi = -\frac{1}{2}((\partial_{x_1}\phi)^2 + (\partial_{x_2}\phi)^2) - g\eta, \tag{5}$$

with g being the gravitational constant. Without loss of generality we will set $g = 1$ and the depth of the fluid at rest, i.e., $\eta = 0$, to one in the following.

Since equation (2) can be solved in $\Omega(t)$, under the boundary condition (3) and $u_1|_{\Gamma(t)} = \partial_{x_1}\phi|_{\Gamma(t)}$ given on the top surface, uniquely up to a constant, all terms on the right hand side of (4) and (5) can be computed if η and $w = u_1|_{\Gamma(t)}$ are known. Hence, the system will be determined by the evolution of the two variables $\eta = \eta(x_1, t)$ and $w = w(x_1, t)$. Therefore, we can restrict ourselves in the following considerations to these variables.

The equations (2)-(5) are called the Eulerian formulation of the water wave problem. We use this relative simple formulation only in the introduction in order to formulate our results. For the proof of the approximation result we will work with the Lagrangian formulation of the water wave problem. For the Eulerian formulation local existence and uniqueness results have been shown for instance in [Sh76, KN79, La05] and for the Lagrangian formulation local existence and uniqueness results have been shown for instance in [Na74, Yo82, Yo83, Cr85, Wu97, Wu99, SW00, Ig01, SW02]. For a third formulation of the water wave problem in which the top surface is parametrized by arc length a local existence and uniqueness theorem has been shown in [Am03, AM05, AM09]. The existence and uniqueness theorems for the water wave problem can be distinguished according to whether or not one considers the 2D or 3D problem, finite or infinite depth, with or without surface tension, regularity of the initial conditions and the coordinates which have been chosen to formulate the problem. Some of these coordinates have the disadvantage of showing secular growth of several variables.

We will derive the NLS equation with the help of the ansatz

$$\begin{pmatrix} \eta \\ w \end{pmatrix} = \varepsilon\Psi_{NLS} + \mathcal{O}(\varepsilon^2)$$

where

$$\varepsilon\Psi_{NLS} = \varepsilon A(\varepsilon(x_1 - c_g t), \varepsilon^2 t) e^{i(k_0 x_1 - \omega_0 t)} \varphi(k_0) + c.c.. \tag{6}$$

Here $0 < \varepsilon \ll 1$ is a small perturbation parameter, $\varphi(k_0) \in \mathbb{C}^2$, c_g being the group velocity of the wave packet and $-\omega_0 < 0$ being the basic temporal wave number associated to the basic spatial wave number $k_0 > 0$. (The minus sign in front on ω_0 simply reflects the fact that we consider right moving waves.) $T = \varepsilon^2 t$ is the slow time scale and $X = \varepsilon(x_1 - c_g t)$ is the slow spatial scale, i.e., the time scale of the modulations is $\mathcal{O}(1/\varepsilon^2)$ and the spatial scale of the modulations is $\mathcal{O}(1/\varepsilon)$. The complex-valued amplitude $A = A(X, T)$ solves in lowest order the NLS equation (1) from the beginning of the paper. By (6) we describe complex-valued slow modulations in time and in space of the underlying temporarily and spatially oscillating wave train $e^{i(k_0 x_1 - \omega_0 t)}$. See Figure 1. The basic spatial wave number $k = k_0$ and the basic temporal wave number $-\omega = -\omega_0$ are related via the linear dispersion relation of the water wave problem (2)-(5), namely

$$L(\omega_\pm, k) = \omega_\pm^2 - k\tanh(k) = 0, \tag{7}$$

where we choose the branch of solutions

$$\omega(k) := \operatorname{sign}(k)\sqrt{k\tanh(k)}\,. \tag{8}$$

Then the group velocity c_g of the wave packet is given by $c_g = \partial_k\omega|_{k=k_0}$. This ansatz leads to waves moving to the right. To obtain waves moving to the left, $-\omega_0$ and c_g have to be replaced by ω_0 and $-c_g$.

It is the purpose of this paper to demonstrate how well the solutions of the 2D water wave problem can be approximated via the formal ansatz (6). As a first step in [CSS92] the so-called residual, i.e., the terms which do not cancel after inserting the ansatz (6) into the equations of the water wave problem (2)-(5) has been estimated in some Sobolev norms. Estimates for the residual in the $3D$-case where the NLS equation is replaced by the Davey-Stewartson system can be found in [CSS97]. The question, if there are solutions of the water wave problem (2)-(5) which behave as predicted by the NLS equation remained open in [CSS92, CSS97]. In [SW11] the NLS approximation has been rigorously justified for a quasilinear reduced model equation for the 2–D water wave problem with finite depth and no surface tension. This reduced model shares with the Lagrangian formulation of the 2–D water wave problem some of the principal difficulties which have to be overcome for a validity proof for the NLS approximation. More recently, Totz and Wu [TW12] have demonstrated the validity of the NLS approximation for the 2–D water wave problem in a channel of infinite depth, though as we explain later in this introduction, we feel that the finite and infinite depth cases are quite different, and techniques applicable in one context do not necessarily transfer to the other.

Notation. We denote the Fourier transform by

$$(\mathcal{F}u)(k) = \hat{u}(k) = \frac{1}{2\pi}\int u(x_1)e^{-ikx_1}dx_1.$$

The Sobolev space H^s is equipped with the norm

$$\|u\|_{H^s} = \left(\int |\hat{u}(k)|^2(1+|k|^2)^s dk\right)^{1/2}.$$

Moreover, let $\|u\|_{C_b^n} = \sum_{j=0}^n \|\partial_x^j u\|_{C_b^0}$, where $\|u\|_{C_b^0} = \sup_{x_1\in\mathbb{R}}|u(x_1)|$.

Because of the loss of smoothness in normal-form transformations we make in a subsequent section, we are forced to work in spaces of analytic functions. Hence, we define

$$Y_{\sigma,s} = \{f \in L^2(\mathbb{R}) \mid \|f\|_{Y_{\sigma,s}} = \left(\int (1+k^2)^s e^{2\sigma|k|}|\hat{f}(k)|^2 dk\right)^{1/2} < \infty\}.$$

Functions in $Y_{\sigma,s}$ are analytic in a strip of width 2σ centered on the real axis.

Our result is

Theorem 1.1. *Fix $s_A \geq s+5 \geq 11$. Then for all $k_0 > 0$ and for all $C_1, T_0 > 0$ there exist $T_1 > 0$, $C_2 > 0$, $\varepsilon_0 > 0$ such that for all solutions $A \in C([0,T_0], H^{s_A}(\mathbb{R},\mathbb{C}))$ of the NLS equation (1) with*

$$\sup_{T\in[0,T_0]} \|A(\cdot,T)\|_{H^{s_A}(\mathbb{R},\mathbb{C})} \leq C_1$$

the following holds. For all $\varepsilon \in (0, \varepsilon_0)$ there exists a solution

$$(\eta, w) \in C([0, T_1/\varepsilon^2], (H^s(I\!\!R, I\!\!R))^2)$$

of the 2D water wave problem (2)-(5) which satisfies

$$\sup_{t \in [0, T_1/\varepsilon^2]} \| \begin{pmatrix} \eta \\ w \end{pmatrix} (\cdot, t) - \varepsilon \Psi_{NLS}(\cdot, t) \|_{(H^s(I\!\!R, I\!\!R))^2} \leq C_2 \varepsilon^{3/2}.$$

The error of order $\mathcal{O}(\varepsilon^{3/2})$ is small compared with the solution (η, w) and the approximation $\varepsilon \Psi_{NLS}$ which are both of order $\mathcal{O}(\varepsilon)$ in L^∞ such that the dynamics of the NLS equation can be found in the water wave problem, too. We note that this fact should not be taken for granted. There are modulation equations (for examples see [Schn95, GS01]) which although derived by reasonable formal arguments do not reflect the true dynamics of the original equations. However, our theorem is not optimal, since in general we cannot prove $T_1 = T_0$. Nevertheless our estimates are on an $\mathcal{O}(1/\varepsilon^2)$ time scale and $T_1 \sim 1/C_1$ has a reasonable size such that the approximation statement is not void.

The NLS equation is a completely integrable Hamiltonian system which can be solved explicitly with the help of some inverse scattering scheme, cf. [AS81]. Our theorem guarantees that for instance parts of the soliton dynamics present in the NLS equation for $\nu_1(k_0)$ and $\nu_2(k_0)$ having the same sign can be found approximately in the water wave problem, too. For a discussion of the values of the coefficients $\nu_j(k_0)$ in (1) see also [AS81, Figure 4.15, p. 321].

The assumption $s \geq 6$ is due to our local existence and uniqueness theory of the water wave problem. The solutions of the NLS equation have to be at least three times more regular than the solutions of the water wave problem due to the fact that the linear dispersion relation (7) has to be expanded at the spatial wave number k_0 up to third order. The additional loss comes mainly from the fact that a higher order approximation is used and that the result is proved for the Lagrangian formulation and then transfered to the Eulerian formulation of the water wave problem.

In order to explain the main ideas of the proof of Theorem 1.1 we consider an abstract evolutionary problem

$$\partial_t v = \Lambda v + B(v, v),$$

with Λ a linear and B a symmetric bilinear operator. Suppose that v is formally approximated by $\varepsilon \Psi_{NLS}$, i.e., that the residual

$$\mathrm{Res}(v) = -\partial_t v + \Lambda v + B(v, v)$$

is small for $v = \varepsilon \Psi_{NLS}$. By modifying the formal approximation $\varepsilon \Psi_{NLS}$ the residual can be made arbitrarily small, i.e., for all $\gamma > 0$ there exists a formal approximation $\varepsilon \Psi$ close to $\varepsilon \Psi_{NLS}$ such that

$$\mathrm{Res}(\varepsilon \Psi) = \mathcal{O}(\varepsilon^\gamma) \qquad \text{and} \qquad \varepsilon \Psi - \varepsilon \Psi_{NLS} = \mathcal{O}(\varepsilon^2). \tag{9}$$

This residual has been estimated in [CSS92]. The estimates contain complicated expansions of the Dirichlet-Neumann operator which appears in the solution of (2).

In order to prove Theorem 1.1 we have to estimate the error

$$\varepsilon^\beta R = v - \varepsilon \Psi$$

for all $t \in [0, T_0/\varepsilon^2]$ to be of order $\mathcal{O}(\varepsilon^\beta)$ for a $\beta > 1$, i.e., we have to prove that R is of order $\mathcal{O}(1)$ for all $t \in [0, T_0/\varepsilon^2]$. The error R satisfies

$$\partial_t R = \Lambda R + 2\varepsilon^\alpha B(\Psi, R) + \varepsilon^\beta B(R, R) + \varepsilon^{-\beta}\mathrm{Res}(\varepsilon\Psi) \ .$$

In our case Λ generates a uniformly bounded semigroup and so we were done beside possible arbitrary complicated functional analytic details, if a) $\alpha \geq 2$, b) $\beta > 2$ and c) $\varepsilon^{-\beta}\mathrm{Res}(\varepsilon\Psi) = \mathcal{O}(\varepsilon^2)$. The result then would follow by a rescaling of time, $T = \varepsilon^2 t$, and an application of Gronwall's inequality (e.g. [KSM92]). In our case, however, we have $\alpha = 1$. We can still make γ in (9) arbitrary large by picking our approximate solution as described below, and in particular, strictly bigger than 4. As a consequence, we can choose $\beta > 2$ and so the points b) and c) are satisfied easily. The difficulty is to control the term $2\varepsilon B(\Psi, R)$ in the linear evolution.

The idea of eliminating this term with a normal-form transform

$$R = w + \varepsilon M(\Psi, w)$$

with M a bilinear mapping goes back to Kalyakin (cf. [Kal88]). See also [Schn98b]. In order to eliminate $2\varepsilon B(\Psi, R)$ by this near identity change of variables a so called non-resonance condition has to be satisfied. The eigenvalues $\lambda_j = \lambda_j(k)$ of the linearized problem (in our problem below, $j = 1, 2, 3$, corresponding to the fact that we write the water wave problem as a system of three equations) as a function over the Fourier wave numbers k have to satisfy

$$|\lambda_p(k) - \lambda_2(k_0) - \lambda_q(k - k_0)| \geq c_{nr} > 0 \ . \tag{10}$$

for $p, q = 1, 2, 3$ and all $k \in I\!R$. It is easy to see that the eigenvalues $\lambda_j = i\omega_j$ of the water wave problem with $\omega_1 = 0$ and $\omega_j = (-1)^{j-1}\omega(k)$ for $j = 2, 3$, where $\omega(k)$ is given by (8), do not satisfy (10) and do possess at least one resonance at the wave number $k = 0$. This resonance is trivial for the water wave problem but a resonance at the wave number $k = 0$ always implies another resonance for the wave number $k = k_0$ which is non-trivial for the water wave problem. A resonance is called trivial if the quadratic terms vanish for the resonant wave number, too. Otherwise it is called non-trivial. For more details, see Section 3. Therefore, [Kal88] is no longer applicable and an improved method developed in [Schn98a] has to be applied. According to an error made in [Schn98a] in the handling of the trivial resonance, the method of [Schn98a] has to be modified slightly, similar to [DS06] and [SW11]. In principle, the method is mainly based on a suitable scaling of the error function R which depends on the wave numbers followed by a number of special normal-form transforms.

After making the normal-form transformation to eliminate the low-order terms in the equation for the remainder, we must control the evolution of the remaining terms. Because the normal-form transformation results in a loss of regularity, the local existence and uniqueness theorems for the water wave problem mentioned above are no longer applicable. Therefore, we proceed as follows. We choose the Lagrangian formulation of the water wave problem and use a Cauchy-Kowalevskaya like method, like [KN79] did for the Eulerian formulation of the water wave problem. We cannot work with the Eulerian formulation since this formulation already loses a derivative on the right hand side which is the maximal allowance for the Cauchy-Kowalevskaya like method. In the Lagrangian formulation the right hand side only loses half a derivative, plus half a derivative from the normal-form transform makes one derivative after the normal-form transform such that the Cauchy-Kowalevskaya method still applies. While this method might seem to require that we consider only analytic solutions of the NLS equation, we can, in fact, consider less smooth solutions, by mollifying them during the approximation process.

Recently, the nature and effects of resonances in the water wave problem has also been examined for the 2D water wave problem by Wu [Wu09] and for the 3D water wave problem in by Germain, Masmoudi and Shatah [GMS12] and with an alternative method by Wu [Wu11] in establishing (almost) global existence results in case of infinite depth, i.e., $\omega^2 = |k|$. However, due to the different goal in [GMS12] the normal-from transformation does not have to be inverted and the loss of regularity occurs in such a way that the local existence method of the untransformed system still can be used.

In case of infinite depth and no surface tension the elimination of all quadratic terms is possible without loss of regularity as has been shown in [Wu09], [Wu11] by using the special structure of this problem. This has been used very recently by Totz and Wu [TW12] to prove the NLS approximation property for the 2D water wave problem in the case of infinite depth and no surface tension. This is the first result establishing the approximation property for the full water wave problem and the NLS equation over the appropriate, NLS, time-scale. However, the methods used in that work are very different from those used in this paper and the differences between the water wave problem in the cases of infinite vs. finite depth are such that a transfer of the results from [TW12] to the case of finite depth does not seem obvious to us.

The justification of the NLS equation in case of positive surface tension will be the subject of further research. For large surface tension or large basic wave number k_0 there are no additional non-trivial resonances. In case of small surface tension additional non-trivial resonances occur and the proof of a possible approximation property will be much more involved. See [Schn05] and [DS06] for the handling of model problems.

The plan of the paper is as follows. In Section 2 we present the Lagrangian formulation of the water wave problem and write it as a first order dynamical system. In a next step the linear part of the Fourier transformed dynamical system is diagonalized. For the diagonalized system we derive the associated NLS equation and construct a modified approximation which makes the residual small. After that we formulate our approximation result for the Lagrangian formulation. In Section 3 we perform the normal-form transform. In this context, special attention is given to the handling of the trivial resonance at the Fourier wavenumber $k = 0$ and of the nontrivial resonance at $k = k_0$. Since the normal-form transform loses regularity it cannot be inverted with the help of a Neumann series. Thus, Section 4 is devoted to the inversion of the normal-form transform. For this purpose we use appropriate energy estimates. In Section 5 we verify that the difference between the true solution of the water wave problem and the (improved) nonlinear Schrödinger approximation remains small over the relevant time scale. We establish the error estimates for the transformed system with the help of the Cauchy-Kowalevskaya like method by proving energy estimates in a scale of Banach spaces of analytic functions.

Further notation and basic facts: We introduce the scaling operator $(S_\varepsilon u)[x_1] = u(\varepsilon x_1)$ and the translation operator $(\tau_y u)[x_1] = u(x_1+y)$. We have $\|S_\varepsilon A\|_{H^m} \leq C\varepsilon^{-1/2}\|A\|_{H^m}$, but $\|S_\varepsilon A\|_{C_b^m} \leq C\|A\|_{C_b^m}$.

In addition to the spaces of analytic functions we introduced earlier we will also sometimes need to use Sobolev spaces with spatial weights, namely the Sobolev space $H^s(m)$ equipped with the norm $\|u\|_{H^s(m)} = \|u\rho^m\|_{H^s}$ where $\rho(x_1) = (1 + x_1^2)^{1/2}$. We also use the notation $L^2(m) = H^0(m)$. We define the space $L^1(m)$ with $u \in L^1(m) \Leftrightarrow u\rho \in L^1$. It is well known that Fourier transform $\mathcal{F}$ is an isomorphism from the space $H^n(m)(\mathbb{R}, \mathbb{C})$ into the space $H^m(n)(\mathbb{R}, \mathbb{C})$. It is a continuous mapping from $L^1(m)$ into C_b^m, but not vice versa. We also define weighted analogues of the analytic function spaces defined above by $Y_{\sigma,s}(m) = \{f \in L^2(I\!R) \mid \rho^m f \in Y_{\sigma,s}\}$.

Throughout this paper many constants are denoted with C - the constant may change without comment in successive inequalities if it can be chosen independently of the spatial perturbation parameter $0 < \varepsilon \ll 1$. The commutator of two operators L and M is defined as $[L, M] = LM - ML$.

Acknowledgments: The work of Guido Schneider was partially supported by the Deutsche Forschungsgemeinschaft DFG under the grant Schn520/3-1/2. The work of Eugene Wayne was supported in part by the NSF through the grant DMS-0908093. Guido Schneider would like to thank the Mathematics Department of Boston University for its hospitality. Eugene Wayne thanks the Mathematics Institute at the University of Stuttgart for its hospitality and the DAAD for support for a visit of one month to Stuttgart.

2 Preparations

2.1 The Lagrangian formulation of the water wave problem

As we mentioned in the introduction the Eulerian formulation (2)-(5) of the 2D water wave problem is not adequate for our purposes. Thus, this subsection is devoted to presenting the Lagrangian formulation of the water wave problem and rewriting this formulation as a first order system of partial differential equations.

For fixed time t the free surface of the fluid can be written as

$$\Gamma(t) = \{(\tilde{X}_1(\alpha, t), \tilde{X}_2(\alpha, t)) = (\alpha + X_1(\alpha, t), X_2(\alpha, t)) | \alpha \in I\!R\} .$$

It is a Jordan-curve which has no intersection with the bottom $\{(\alpha, -1)|\alpha \in I\!R\}$. Under the assumptions on the flow which we made in the introduction the dynamics of the 2D water wave problem is completely determined by the evolution of the free surface $\Gamma(t)$ which is governed by (for a careful derivation of the following system of equations see [Yo82])

$$\begin{aligned} \partial_t^2 X_1(1 + \partial_\alpha X_1) + \partial_\alpha X_2(1 + \partial_t^2 X_2) &= 0, && (11) \\ \partial_t X_2 &= \mathcal{K}(X)\partial_t X_1. && (12) \end{aligned}$$

The operator $\mathcal{K}(X)$ depends linearly on $U_1 = \partial_t X_1$, but nonlinearly on X. It is related to the Dirichlet-Neumann operator and its existence is a consequence of the incompressibility and irrotationality of the flow. It is defined by $\mathcal{K}(X)U_1 = \partial_{x_2}\phi|_{\Gamma(t)}$, where $\phi : \Omega(t) \to I\!R$ solves for fixed t the boundary value problem

$$\begin{aligned} \Delta\phi = 0, &\quad \text{in} \quad \Omega(t), \\ \partial_{x_2}\phi = 0, &\quad \text{for} \quad x_2 = -1, \\ \partial_{x_1}\phi = U_1, &\quad \text{on} \quad \Gamma(t). \end{aligned}$$

The operator $\mathcal{K}(X)$ is of the form $\mathcal{K}(X) = \mathcal{K}_0 + \mathcal{S}_1(X)$, where $\mathcal{K}_0$ is the linear part of the operator $\mathcal{K}(X)$, and has the Fourier symbol $\hat{\mathcal{K}}_0(k) = -i\tanh(k)$. The nonlinear part $\mathcal{S}_1(X)\cdot$ has certain smoothing properties which are summarized in Appendix A.1. In particular, we prove that $\mathcal{K}(X)$ (and as a consequence, $\mathcal{S}_1(X)$) depends analytically on $\partial_\alpha X_1, X_2 \in Y_{\sigma,s}$ for any $\sigma > 0, s > 1$.

As explained in [SW00], due to the behavior of the system at the wave number $k = 0$ the variable X_1 is unbounded in space and grows rapidly in time for the approximation which makes the resulting solutions difficult to control over the long time scales which we need to work with. However, as we also discussed in that reference, the derivatives of X_1 do not suffer from this secular growth, and thus it is advantageous to work with the variable $Z_1 = \mathcal{K}_0 X_1$ (which for "long-wavelength" initial conditions behaves like $Z_1 \approx -\partial_\alpha X_1$.) Somewhat surprisingly the system of equations for the water wave problem can be rewritten entirely in terms of the variables $\mathcal{W} = (Z_1, X_2, U_1)$, namely

$$\partial_t \mathcal{W} = F_{\mathcal{W}}(\mathcal{W}) \tag{13}$$

with

$$F_{\mathcal{W}}(\mathcal{W}) = \begin{pmatrix} \mathcal{K}_0 U_1 \\ \mathcal{K}_0 U_1 + \mathcal{S}_1(X) U_1 \\ -(1 - \mathcal{M}_2 Z_1 + (\partial_\alpha X_2)\mathcal{K}_0 + (\partial_\alpha X_2)\mathcal{S}_1(X))^{-1}[(\partial_\alpha X_2)(1 + [\partial_t, \mathcal{S}_1(X)]U_1)] \end{pmatrix},$$

where

$$\mathcal{M}_2 \cdot = -\partial_\alpha (\mathcal{K}_0)^{-1} \cdot \ .$$

From the estimates on $\mathcal{K}(X)$ proved in Appendix A.1 we see that $F_{\mathcal{W}}$ is an analytic mapping from $Y_{\sigma,s} \times Y_{\sigma,s} \times Y_{\sigma,s-\frac{1}{2}}$ into $Y_{\sigma,s-\frac{1}{2}} \times Y_{\sigma,s-\frac{1}{2}} \times Y_{\sigma,s-1}$

Moreover we define the vector $\mathcal{V} = (X_1, X_2, U_1)$. We also abuse notation slightly and do not distinguish between operators which depend on $\mathcal{V}$ or $\mathcal{W}$, i.e., for instance we will write $\mathcal{K}(X)$ as either $\mathcal{K}(\mathcal{V})$ or $\mathcal{K}(\mathcal{W})$, depending on the circumstances.

Remark 2.1. Note that from (11)-(12) to (13) information is lost. In order to compute the physical solution the point $X_1(0,t)$ has to be computed a posteriori from $X_1(0,0)$ and $U_1(0,t)$ which is contained in (13). However, (13) is independent of $X_1(0,t)$. This corresponds to the fact that the bottom of the canal can be shifted without having any influence on the dynamics in $\Omega(t)$.

Remark 2.2. The choice of variables (Z_1, X_2, U_1) has the additional advantage that all variables will scale the same in terms of the small perturbation parameter for $\varepsilon \to 0$. (X_1, X_2, U_1) scale differently at the wave number $k = 0$.

For completeness we close this section with some remarks about the existence and uniqueness of solutions of system (13) which is obtained indirectly. System (13) is embedded in a larger quasilinear system of PDE's for which standard local existence and uniqueness techniques apply.

From [SW00] we have the following local and uniqueness theorem.

Theorem 2.3. *Define the space $\mathcal{H}^s = H^s \times H^s \times H^{s-1/2}$. For all $s \geq 6$ there exists a $C_1 > 0$ such that for all $C_2 \in (0, C_1]$ we have a $t_0 > 0$ such that the following is true. For each initial condition $\mathcal{W}_0 \in \mathcal{H}^s$ with $\|\mathcal{W}_0\|_{\mathcal{H}^s} \leq C_2$ there exists a unique solution $\mathcal{W} \in C([0,t_0], \mathcal{H}^s)$ of (13) with $\mathcal{W}|_{t=0} = \mathcal{W}_0$.*

However, as explained in the introduction, due to a loss of smoothness in the normal-form transformations this existence theory is insufficient to prove the accuracy of the approximation by the NLS equation and hence in Section 5 we prove a new existence theorem in the analytic function spaces introduced above.

2.2 The diagonalization

In order to construct the normal-form transformations it is useful to diagonalize the linear part of (13). In Fourier space the linearization is given by

$$\partial_t \begin{pmatrix} \hat{Z}_1 \\ \hat{X}_2 \\ \hat{U}_1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -i\tanh(k) \\ 0 & 0 & -i\tanh(k) \\ 0 & -ik & 0 \end{pmatrix} \begin{pmatrix} \hat{Z}_1 \\ \hat{X}_2 \\ \hat{U}_1 \end{pmatrix}.$$

The eigenvalues of the matrix on the right hand side are given by $\lambda_j = i\omega_j$ for $j = 1, 2, 3$ with

$$\omega_1(k) = 0, \qquad \omega_2(k) = -\omega(k), \qquad \omega_3(k) = \omega(k),$$

where

$$\omega(k) = \operatorname{sign}(k)\sqrt{k\tanh(k)}.$$

We write the original coordinates as sum of the associated eigenvectors, i.e.,

$$\begin{aligned} \begin{pmatrix} \hat{Z}_1 \\ \hat{X}_2 \\ \hat{U}_1 \end{pmatrix} &= \hat{c}_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \hat{c}_2 \begin{pmatrix} \hat{s} \\ \hat{s} \\ 1 \end{pmatrix} + \hat{c}_3 \begin{pmatrix} -\hat{s} \\ -\hat{s} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & \hat{s} & -\hat{s} \\ 0 & \hat{s} & -\hat{s} \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \end{pmatrix} = \mathcal{D}(k) \begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \end{pmatrix}, \end{aligned}$$

with $\hat{s} = \hat{s}(k) = \sqrt{k^{-1}\tanh(k)}$. The adjoint eigenvectors are given

$$\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ 1/(2\hat{s}) \\ 1/2 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ -1/(2\hat{s}) \\ 1/2 \end{pmatrix}.$$

Due to the asymptotic behavior of $\hat{s}$ it is easy to see that from $(Z_1, X_2, U_1) \in Y_{\sigma,s} \times Y_{\sigma,s} \times Y_{\sigma,s-\frac{1}{2}}$, it follows $(c_1, c_2, c_3) \in Y_{\sigma,s} \times Y_{\sigma,s-\frac{1}{2}} \times Y_{\sigma,s-\frac{1}{2}}$ and vice versa. The variables $c = (c_1, c_2, c_3)^T$ satisfy

$$\partial_t c = F_c(c) = \mathcal{D}^{-1} F_{\mathcal{W}}(\mathcal{D}c) \tag{14}$$

For the same reason, F_c is a smooth mapping from $H^s \times H^{s-1/2} \times H^{s-1/2}$ into $H^{s-1/2} \times H^{s-1} \times H^{s-1}$.

According to the fact that the quadratic terms play the major role in the following we expand (13) up to terms of quadratic order and find with [SW00, Lemma 3.8] and [SW00, Remark 3.9] based on [Cr85, Lemma 3.7: page 827] that

$$\begin{aligned} \partial_t Z_1 &= \mathcal{K}_0 U_1, \\ \partial_t X_2 &= \mathcal{K}_0 U_1 + \mathcal{M}_1(Z_1, \partial_\alpha U_1) - (X_2 + \mathcal{K}_0(X_2\mathcal{K}_0))\partial_\alpha U_1 + \mathcal{O}(\|\mathcal{W}\|^3), \\ \partial_t U_1 &= -\partial_\alpha X_2 - (\mathcal{M}_2 Z_1)\partial_\alpha X_2 + (\partial_\alpha X_2)\mathcal{K}_0 \partial_\alpha X_2 + \mathcal{O}(\|\mathcal{W}\|^3), \end{aligned} \tag{15}$$

where $\mathcal{M}_1(Z_1, \cdot) = [X_1, \mathcal{K}_0]\cdot$. Here, $\|\mathcal{W}\| = \|Z_1\|_{Y_{\sigma,s}} + \|X_2\|_{Y_{\sigma,s}} + \|U_1\|_{Y_{\sigma,s-\frac{1}{2}}}$. The notation $\mathcal{O}(\|\mathcal{W}\|^3)$ means that the terms omitted from the equation can be bounded by $C\|\mathcal{W}\|^3$, an estimate which follows from the analyticity of the nonlinearity in (13).

The system for c_1, c_2, c_3 is then given by

$$
\begin{aligned}
\partial_t c_1 &= -\mathcal{M}_1(c_1 + sc_2 - sc_3, \partial_\alpha(c_2 + c_3)) \\
&\quad +((sc_2 - sc_3) + \mathcal{K}_0((sc_2 - sc_3)\mathcal{K}_0))\partial_\alpha(c_2 + c_3) + \mathcal{O}(\|c\|^3), \\
\partial_t c_2 &= -i\omega c_2 + \frac{1}{2s}(\mathcal{M}_1(c_1 + sc_2 - sc_3, \partial_\alpha(c_2 + c_3)) \\
&\quad -((sc_2 - sc_3) + \mathcal{K}_0((sc_2 - sc_3)\mathcal{K}_0))\partial_\alpha(c_2 + c_3)) \\
&\quad -\frac{1}{2}(\mathcal{M}_2(c_1 + sc_2 - sc_3))\partial_\alpha(sc_2 - sc_3) \\
&\quad +\frac{1}{2}(\partial_\alpha(sc_2 - sc_3))\mathcal{K}_0\partial_\alpha(sc_2 - sc_3) + \mathcal{O}(\|c\|^3), \\
\partial_t c_3 &= i\omega c_3 - \frac{1}{2s}(\mathcal{M}_1(c_1 + sc_2 - sc_3, \partial_\alpha(c_2 + c_3)) \\
&\quad -((sc_2 - sc_3) + \mathcal{K}_0((sc_2 - sc_3)\mathcal{K}_0))\partial_\alpha(c_2 + c_3)) \\
&\quad -\frac{1}{2}(\mathcal{M}_2(c_1 + sc_2 - sc_3))\partial_\alpha(sc_2 - sc_3) \\
&\quad +\frac{1}{2}(\partial_\alpha(sc_2 - sc_3))\mathcal{K}_0\partial_\alpha(sc_2 - sc_3) + \mathcal{O}(\|c\|^3),
\end{aligned}
\tag{16}
$$

where $\|c\| = \|c_1\|_{Y_{\sigma,s}} + \|c_2\|_{Y_{\sigma,s-\frac{1}{2}}} + \|c_3\|_{Y_{\sigma,s-\frac{1}{2}}}$.

2.3 Derivation of the NLS equation

In order to derive the NLS equation we make the ansatz

$$
\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \varepsilon\Psi_1 + \varepsilon\Psi_{-1} + \varepsilon^2\Psi_0 + \varepsilon^2\Psi_2 + \varepsilon^2\Psi_{-2} \tag{17}
$$

with

$$
\begin{aligned}
\varepsilon\Psi_{\pm1} &= \varepsilon\psi_{\pm1}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \varepsilon A_{\pm1}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\,\mathbf{E}^{\pm1}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\
\varepsilon^2\Psi_0 &= \begin{pmatrix} \varepsilon^2\psi_{01} \\ \varepsilon^2\psi_{02} \\ \varepsilon^2\psi_{03} \end{pmatrix} = \begin{pmatrix} \varepsilon^2 A_{01}(\varepsilon(\alpha - c_g t), \varepsilon^2 t) \\ \varepsilon^2 A_{02}(\varepsilon(\alpha - c_g t), \varepsilon^2 t) \\ \varepsilon^2 A_{03}(\varepsilon(\alpha - c_g t), \varepsilon^2 t) \end{pmatrix}, \\
\varepsilon^2\Psi_{\pm2} &= \begin{pmatrix} \varepsilon^2\psi_{(\pm2)1} \\ \varepsilon^2\psi_{(\pm2)2} \\ \varepsilon^2\psi_{(\pm2)3} \end{pmatrix} = \begin{pmatrix} \varepsilon^2 A_{(\pm2)1}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\,\mathbf{E}^{\pm2} \\ \varepsilon^2 A_{(\pm2)2}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\,\mathbf{E}^{\pm2} \\ \varepsilon^2 A_{(\pm2)3}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\,\mathbf{E}^{\pm2} \end{pmatrix},
\end{aligned}
$$

where $\mathbf{E} = e^{i(k_0\alpha - \omega_0 t)}$, $\omega_0 = \omega(k_0)$, $\overline{A}_j = A_{-j}$ and $\overline{A}_{jl} = A_{-jl}$. Here, and throughout the remainder of the paper, we will use upper case Ψ to denote vector valued functions and lower case ψ to denote scalar functions.

Remark 2.4. *Our ansatz leads to waves moving to the right. For waves moving to the left one has to replace in the above ansatz the vector* $(0,1,0)^T$ *by* $(0,0,1)^T$ *as well as* $-\omega_0$ *by* ω_0 *and* c_g *by* $-c_g$.

We equate the coefficients of the $\varepsilon^m \mathbf{E}^j$ to zero and find that the coefficients of $\varepsilon \mathbf{E}^1$ and $\varepsilon^2 \mathbf{E}^1$ vanish identically due to the definition of $\omega = \omega(k)$ and $c_g = c_g(k)$. For $\varepsilon^3 \mathbf{E}^1$ we obtain

$$\partial_T A_1 = -\frac{i\omega''(k_0)}{2} \partial_X^2 A_1 + \text{ nonlinear terms.}$$

The nonlinear terms are a sum of multiples of $A_1|A_1|^2$, $A_1 A_{0l}$, and $A_{-1}A_{2l}$. In the next steps we obtain algebraic relations such that the A_{2l} can be expressed in terms of A_1^2 and the A_{0l} in terms of $|A_1|^2$, respectively.

For $\varepsilon^2 \mathbf{E}^2$ we obtain

$$\begin{aligned} -2\omega_0 A_{21} &= \gamma_{21} A_1^2 \\ (-2\omega_0 + \omega(2k_0))A_{22} &= \gamma_{22} A_1^2, \\ (-2\omega_0 - \omega(2k_0))A_{23} &= \gamma_{23} A_1^2 \end{aligned} \tag{18}$$

with coefficients $\gamma_{2l} \in \mathbb{C}$. Since $2\omega_0 \neq 0$, $-2\omega_0 + \omega(2k_0) \neq 0$ and $-2\omega_0 - \omega(2k_0) \neq 0$, which follows from the explicit form of $\omega(k)$, the A_{2l} are well-defined in terms of A_1^2.

All terms vanish identically for $\varepsilon^2 \mathbf{E}^0$. This is obvious for the linear terms. For the quadratic terms the calculations are analogous to those of Appendix A of [SW11] (see specifically equation (94)). The nonlinear terms in $\varepsilon^3 \mathbf{E}^0$ must be proportional to ∂_X since no other combination of terms in the approximation (17) leads to terms proportional to $\varepsilon^3 \mathbf{E}^0$. So we find

$$\begin{aligned} 0 &= -c_g \partial_X A_{01} + \gamma_{01} \partial_X (A_1 A_{-1}), \\ -c_g \partial_X A_{02} &= -\omega'(0) \partial_X A_{02} + \gamma_{02} \partial_X (A_1 A_{-1}), \\ -c_g \partial_X A_{03} &= \omega'(0) \partial_X A_{03} + \gamma_{03} \partial_X (A_1 A_{-1}), \end{aligned}$$

where now $\gamma_{0l} \in \mathbb{R}$ according to the fact that we consider a real-valued problem. Since $c_g \notin \{0, -\omega'(0), \omega'(0)\}$ we can divide the equations for $\varepsilon^3 \mathbf{E}^0$ by ∂_X and can express the A_{0l} in terms of $|A_1|^2$.

As mentioned above the nonlinear terms in the equation for $\varepsilon^3 \mathbf{E}^1$ include $A_1|A_1|^2$ as well as terms consisting of combinations of A_1 with the A_{0l} and of A_{-1} with the A_{2l}. Eliminating A_{0l} and A_{2l} by the algebraic relations obtained for $\varepsilon^3 \mathbf{E}^0$ and $\varepsilon^2 \mathbf{E}^2$ gives finally the NLS equation

$$\partial_T A_1 = -i \frac{\omega''(k_0)}{2} \partial_X^2 A_1 + i\nu_2(k_0) A_1 |A_1|^2$$

with a $\nu_2(k_0) \in \mathbb{R}$.

2.4 The modified approximation

After the derivation of the NLS equation in the last section, this section is devoted to the construction of the improved approximation $\varepsilon\Psi$. We proceed in two steps. First, the above approximation is extended by higher order terms. Secondly, the support of the modified approximation in Fourier space is restricted to small neighborhoods of integer multiples of the basic wave number $k_0 > 0$ by introducing some cut-off functions. Since the approximation in Fourier space is strongly concentrated around these wave numbers the approximation is only changed slightly by this modification, but this second step will give us a much simpler control on the resonances and makes $\varepsilon\Psi$ an analytic function.

As explained in the introduction the approximation (17) is modified in order to make the so called residual

$$\mathrm{Res}(c) = -\partial_t c + F_c(c)$$

small. The residual will contain all terms which do not cancel after inserting the approximation into the equations. Hence, $\mathrm{Res}(c) = 0$ if and only if c solves (14). In a first step the above approximation is extended by higher order terms. Therefore, we define

$$\begin{aligned}
\varepsilon^3 \Psi_h \quad = \quad & \sum_{j=-1,1} \begin{pmatrix} \varepsilon^3 A^1_{j1}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^3 A^1_{j2}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^3 A^1_{j3}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \end{pmatrix} \\
& + \sum_{j=-3,3} \begin{pmatrix} \varepsilon^3 A_{j1}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^3 A_{j2}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^3 A_{j3}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \end{pmatrix} \\
& + \begin{pmatrix} \varepsilon^4 A^1_{01}(\varepsilon(\alpha - c_g t), \varepsilon^2 t) \\ \varepsilon^4 A^1_{02}(\varepsilon(\alpha - c_g t), \varepsilon^2 t) \\ \varepsilon^4 A^1_{03}(\varepsilon(\alpha - c_g t), \varepsilon^2 t) \end{pmatrix} \\
& + \sum_{j=-2,2} \begin{pmatrix} \varepsilon^4 A^1_{j1}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^4 A^1_{j2}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^4 A^1_{j3}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \end{pmatrix} \\
& + \sum_{j=-4,4} \begin{pmatrix} \varepsilon^4 A_{j1}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^4 A_{j2}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \\ \varepsilon^4 A_{j3}(\varepsilon(\alpha - c_g t), \varepsilon^2 t)\mathbf{E}^j \end{pmatrix}.
\end{aligned}$$

Inserting

$$\varepsilon\Psi = \varepsilon\Psi_1 + \varepsilon\Psi_{-1} + \varepsilon^2\Psi_q$$

with

$$\varepsilon^2\Psi_q = \begin{pmatrix} \varepsilon^2\psi_{q1} \\ \varepsilon^2\psi_{q2} \\ \varepsilon^2\psi_{q3} \end{pmatrix} = \varepsilon^2\Psi_0 + \varepsilon^2\Psi_2 + \varepsilon^2\Psi_{-2} + \varepsilon^3\Psi_h$$

into the equations of the water wave problem will show that the residual is formally at least of order $\mathcal{O}(\varepsilon^5)$ if the A_{jl}, A^1_{jl} are chosen in a suitable way. Again we equate the coefficients of the $\varepsilon^m\mathbf{E}^j$ to zero. The resulting equations

$$\begin{aligned}
-j\omega_0 A_{j1} \quad &= \quad \text{nonlinear terms}, \\
(-j\omega_0 + \omega(jk_0))A_{j2} \quad &= \quad \text{nonlinear terms}, \\
(-j\omega_0 - \omega(jk_0))A_{j3} \quad &= \quad \text{nonlinear terms},
\end{aligned} \tag{19}$$

for $j \in \{3, 4\}$ and

$$\begin{aligned}
-2\omega_0 A^1_{21} \quad &= \quad \text{nonlinear terms}, \\
(-2\omega_0 + \omega(2k_0))A^1_{22} \quad &= \quad \text{nonlinear terms}, \\
(-2\omega_0 - \omega(2k_0))A^1_{23} \quad &= \quad \text{nonlinear terms},
\end{aligned} \tag{20}$$

for $j = 2$ can be resolved with respect to the A_{jl}, A^1_{2l} since we have the validity of the non-resonance conditions $j\omega_0 \neq 0$, $-j\omega_0 + \omega(jk_0) \neq 0$, and $-j\omega_0 - \omega(jk_0) \neq 0$ for all $j \in \{2, 3, 4\}$.

For $j=1$ we obtain

$$\begin{aligned}
-\omega_0 A_{11}^1 &= \text{ nonlinear terms,} \\
\partial_T A_{12}^1 &= -\frac{i\omega''(k_0)}{2}\partial_X^2 A_{12}^1 + \text{ nonlinear terms,} \\
(-\omega_0-\omega(k_0))A_{13}^1 &= \text{ nonlinear terms,}
\end{aligned} \tag{21}$$

where the nonlinear terms in the second equation depend linearly on A_{12}^1. The first and the third equation can be resolved with respect to A_{11}^1 and A_{13}^1 since $\omega_0 \neq 0$ and $-\omega_0-\omega(k_0) \neq 0$, respectively.

For $j=0$ we obtain

$$\begin{aligned}
0 &= -c_g\partial_X A_{01}^1 + \text{ nonlinear terms,} \\
-c_g\partial_X A_{02}^1 &= -\omega'(0)\partial_X A_{02}^1 + \text{ nonlinear terms,} \\
-c_g\partial_X A_{03}^1 &= \omega'(0)\partial_X A_{03}^1 + \text{ nonlinear terms.}
\end{aligned} \tag{22}$$

The equations can be solved for A_{0l}^1 since all nonlinear terms are of the form of some previously determined expression, differentiated with respect to X. Since $\omega'(0) \neq \pm c_g$, we can determine A_{0l}^1 by a straightforward integration.

We have the following scheme ($l=1,2,3,\ d=1,3$). The first group of equations is given by

$$\partial_T A_1 = -i\frac{\omega''(k_0)}{2}\partial_X^2 A_1 + i\nu_2(k_0)A_1|A_1|^2, \tag{23}$$

$$A_{2l} = \ell_{2l}(A_1A_1), \tag{24}$$

$$A_{0l} = \ell_{0l}(A_1A_{-1}), \tag{25}$$

where $\nu_2(k_0) \in \mathbb{R}$ and the ℓs are linear maps in their arguments, which can be computed as discussed above. The second group of equations is given by

$$\begin{aligned}
\partial_T A_{12}^1 &= -i\frac{\omega''(k_0)}{2}\partial_X^2 A_{12}^1 \\
&\quad + \ell_{12}^1(A_{0l}^1A_1, A_{2l}^1A_{-1}, A_{0l}A_{12}^1, A_{2l}A_{-12}^1, A_{12}^1A_1A_{-1}, A_1A_1A_{-12}^1) + f_{12}^1,
\end{aligned} \tag{26}$$

$$A_{1d}^1 = f_{1d}^1, \tag{27}$$

$$A_{3l} = f_{3l}, \tag{28}$$

where the ℓs are linear maps in their arguments and the fs functions of the variables of the first group. The third group of equations is given by

$$A_{2l}^1 = \ell_{2l}^1(A_{12}^1A_1) + f_{2l}^1, \tag{29}$$

$$A_{0l}^1 = \ell_{0l}^1(A_{12}^1A_{-1}, A_1A_{-12}^1) + f_{0l}^1, \tag{30}$$

$$A_{4l} = f_{4l}, \tag{31}$$

where the ℓs are linear maps in their arguments and the fs functions of the variables of the first and second group. Moreover, we have the relations

$$A_{-1} = \overline{A}_1, \tag{32}$$

$$A_{-jl} = \overline{A}_{jl}, \tag{33}$$

$$A_{-jl}^1 = \overline{A}_{jl}^1. \tag{34}$$

The NLS equation is the only nonlinear equation. All other equations are linear partial differential equations, or linear algebraic equations. Therefore, we have

Theorem 2.5. *Fix $s_A \geq s+5 \geq 11$. Let $A_1 \in C([0,T_0], H^{s_A})$ be a solution of the Nonlinear Schrödinger equation (23). Then the A_j, A_{jl} and A_{jl}^1 defined through (24)-(34) exist for all $t \in [0,T_0]$ and satisfy $A_j, A_{jl}, A_{jl}^1 \in C([0,T_0], H^{s_A})$.*

From the form of the Ansatz for $\varepsilon\Psi$ and the discussion above we see that formally the residual is now of order $\mathcal{O}(\varepsilon^5)$. Furthermore, the nonlinear terms in the water wave problem all contain a derivative (or a factor of $\mathcal{K}_0$, or a commutator, both of which behave like a derivative for wavenumbers near $k=0$.) If we consider the behavior of the residual near $k=0$, we see that the contribution comes from terms in which all factors of $\mathbf{E}^{\pm j}$ have cancelled. Thus, the derivative must act on a factor of $A_{k\ell}^j$, and because of the long-wave character of these terms, the derivative creates an additional power of ε - i.e. the residual is actually of $\mathcal{O}(\varepsilon^6)$ near $k=0$. Alternatively, one could extend the approximation $\varepsilon\Psi$ by terms of even higher order, as was done in [SW11], resulting in a residual of order $\mathcal{O}(\varepsilon^m)$ with $m \geq 6$.

For the subsequent analysis it is advantageous to modify Ψ further. With the help of the characteristic function

$$\chi_{[-\delta,\delta]}(k) = \begin{cases} 1, & k \in [-\delta,\delta], \\ 0, & k \notin [-\delta,\delta] \end{cases}$$

we introduce the operator $(E_\delta u)(\alpha) = \mathcal{F}^{-1}(\chi_{[-\delta,\delta]}\mathcal{F}u)(\alpha)$. It allows us to extract the Fourier modes belonging to intervals of wave numbers. According to the existing literature the operators E_δ are called mode filters. Such mode filters work as follows:

$$\begin{aligned} \|E_\delta S_\varepsilon u - S_\varepsilon u\|_{H^s} &\leq C\|(\chi_{[-\delta,\delta]}-1)\varepsilon^{-1}S_{1/\varepsilon}\hat{u}\|_{H^0(s)} \\ &\leq \sup_{k\in\mathbb{R}}|\frac{(\chi_{[-\delta,\delta]}-1)(1+k^2)^{s/2}}{(1+|k/\varepsilon|^2)^{(s+5)/2}}|\,\varepsilon^{-1/2}\|u\|_{H^{s+5}} \\ &\leq C\varepsilon^{s+9/2}\|u\|_{H^{s+5}}. \end{aligned} \tag{35}$$

Thus, we set $\delta = k_0/16$ and modify our approximation by replacing all terms of the form $S_\varepsilon\tau_{c_gt}A_j$, $S_\varepsilon\tau_{c_gt}A_{jl}$ and $S_\varepsilon\tau_{c_gt}A_{jl}^n$ in the approximation Ψ by

$$S_\varepsilon\tau_{c_gt}\tilde{A}_j := E_\delta S_\varepsilon\tau_{c_gt}A_j, \tag{36}$$

$$S_\varepsilon\tau_{c_gt}\tilde{A}_{jl} := E_\delta S_\varepsilon\tau_{c_gt}A_{jl}, \tag{37}$$

$$S_\varepsilon\tau_{c_gt}\tilde{A}_{jl}^1 := E_\delta S_\varepsilon\tau_{c_gt}A_{jl}^1. \tag{38}$$

(We note that heretofore, we have written terms like $S_\varepsilon\tau_{c_gt}A_j$ as $\varepsilon A_j(\varepsilon(\alpha-c_gt),\varepsilon^2 t))$.

Then, for the residual the following estimates hold.

Lemma 2.6. *Let $s_A \geq s+5 \geq 11$ and $A_1 \in C([0,T_0], H^{s_A}(\mathbb{R},\mathbb{C}))$ be a solution of the NLS equation (23). Assume that the $\tilde{A}_j$, $\tilde{A}_{jl}$, $\tilde{A}_{jl}^1$ satisfy (36)-(38), where the A_j, A_{jl}, A_{jl}^1 solve (23)-(34). Then there exist $C_{Res}, \varepsilon_0 > 0$ such that for all $\varepsilon \in (0,\varepsilon_0)$ the corresponding approximation $\varepsilon\Psi$ satisfies*

$$\begin{aligned} \sup_{t\in[0,T_0/\varepsilon^2]} \|\mathrm{Res}(\varepsilon\Psi)\|_{H^s} &\leq C_{Res}\varepsilon^{9/2}, \\ \sup_{t\in[0,T_0/\varepsilon^2]} \|E_\delta\mathrm{Res}(\varepsilon\Psi)\|_{H^s} &\leq C_{Res}\varepsilon^{11/2}, \\ \sup_{t\in[0,T_0/\varepsilon^2]} \|\varepsilon\Psi - \varepsilon\Psi_{NLS}\|_{H^s} &\leq C_{Res}\varepsilon^{3/2}. \end{aligned}$$

Moreover, we have

$$\sup_{t\in[0,T_0/\varepsilon^2]} (\|\varepsilon\hat{\Psi}_{\pm1}\|_{L^1(s+1)} + \|\varepsilon\hat{\Psi}_q\|_{L^1(s+1)}) \quad = \quad \mathcal{O}(1).$$

Proof. Due to our "cut-off"-procedure the residual can be written as

$$\mathrm{Res}(\Psi) = \sum_{j=-4}^{4} a_j \text{ with } \mathrm{supp}\mathcal{F}a_j \subset [jk_0 - k_0/4, jk_0 + k_0/4].$$

By construction the above results hold since we have additionally the estimate (35). Therefore, we are done. □

Remark 2.7. The bound on $C_{\mathrm{Res}}\varepsilon^{9/2}$ rather than $C\varepsilon^5$ on the residual is simply a result of the way the L^2-norms scale, i.e., if $A \in L^2$, and $(\mathcal{S}_\varepsilon A)(x) = A(\varepsilon x)$, then $\|\mathcal{S}_\varepsilon A\|_{L^2} = \varepsilon^{-1/2}\|A\|_{L^2}$, cf. (35). In contrast we have $\|u\|_{C_b^0} = \|\mathcal{S}_\varepsilon u\|_{C_b^0}$ and since $\mathcal{F}(\mathcal{S}_\varepsilon A) = \varepsilon^{-1}\mathcal{S}_{1/\varepsilon}(\mathcal{F}A)$ we have $\|\hat{u}\|_{L^1} = \|\varepsilon^{-1}\mathcal{S}_{1/\varepsilon}\hat{u}\|_{L^1}$. The last estimates are used for instance to estimate

$$\|\Psi R\|_{H^s} \leq C\|\Psi\|_{C_b^s}\|R\|_{H^s} \leq C\|\hat{\Psi}\|_{L^1(s)}\|R\|_{H^s}$$

without loss of powers in ε.

2.5 The approximation result in Lagrangian coordinates

Our result for the water wave problem in Lagrangian coordinates is as follows

Theorem 2.8. *Fix $s_A \geq s+5 \geq 11$, let $\beta = 7/2$, and let $\mathcal{H}^s = H^s \times H^s \times H^{s-1/2}$. For all C_A, C_0, $T_0 > 0$ there exist C_R, ε_0, $T_1 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ the following is true. Let $A \in C([0,T_0], H^{s_A})$ be a solution of (1) with*

$$\sup_{T\in[0,T_0]} \|A\|_{H^{s_A}} \leq C_A$$

and let $\mathcal{W}|_{t=0} = \varepsilon\Psi|_{t=0} + \varepsilon^\beta R|_{t=0} \in \mathcal{H}^s$ with $\|R|_{t=0}\|_{\mathcal{H}^s} \leq C_0$. Then there is a unique solution $\mathcal{W} = \varepsilon\Psi + \varepsilon^\beta R \in C([0,T_1/\varepsilon^2], \mathcal{H}^s)$ of (13) which satisfies

$$\sup_{t\in[0,T_1/\varepsilon^2]} \|R(t)\|_{\mathcal{H}^s} \leq C_R.$$

Before we start to prove this approximation result, we show how it relates to the formulation in the introduction.

Proof of Theorem 1.1: Theorem 1.1 is a consequence of Theorem 2.8. The estimates for the Eulerian variables $w = w(x,t)$ and $\eta = \eta(x,t)$ defined by

$$w(\tilde{X}_1(\alpha,t),t) = \partial_t X_1(\alpha,t) \qquad \text{and} \qquad \eta(\tilde{X}_1(\alpha,t),t) = X_2(\alpha,t)$$

follow in a fashion very similar to that of [SW03]. Let $A \in C([0,T_0], H^{s_A})$ be a solution of (1) and construct $\varepsilon\Psi$ as in Subsection 2.4. Let $\mathcal{W} = \varepsilon\Psi_{\mathcal{W}} + \varepsilon^\beta R_{\mathcal{W}}$ with $\Psi_{\mathcal{W}} = (\Psi_{Z_1}, \Psi_{X_2}, \Psi_{U_1}) = \mathcal{D}\Psi$ and $R_{\mathcal{W}} = (R_{Z_1}, R_{X_2}, R_{U_1}) = \mathcal{D}R$ be the solution of (13) constructed in Theorem 2.8. Note that $X_1(\alpha,t) = X_1(\alpha,0) + \int_0^t U_1(\alpha,\tau)d\tau$. Theorem 2.8 implies that $U_1 = \varepsilon\Psi_{U_1} + \varepsilon^\beta R_{U_1}$, with $\beta = 7/2$. We find with functions $A^j \in C([0,T_0], H^{s_A})$ such that

$$\int_0^t U_1(\alpha,\tau)d\tau = \sum_{j=-3}^{3} \varepsilon^{||j|-1|+1}(-i\omega_0)\int_0^t A^j(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)\mathbf{E}^j d\tau + c.c. + \int_0^t U_1^{rem}(\alpha,\tau)d\tau$$

where $\| \int_0^t U_1^{rem}(\cdot,\tau)d\tau\|_{H^s} \leq Ct\varepsilon^{7/2}$. Turning our attention to the integral involving A^1 we see that

$$\begin{aligned}
\int_0^t A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}d\tau &= -\int_0^t A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)\frac{1}{i\omega_0}\partial_\tau(e^{i(k_0\alpha-\omega_0\tau)})d\tau \\
&= -\frac{1}{i\omega_0}A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}|_0^t \qquad (39)\\
&\quad +\frac{1}{i\omega_0}\varepsilon^2\int_0^t \partial_T A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}d\tau \\
&\quad +\frac{1}{i\omega_0}\varepsilon\int_0^t \partial_X A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}d\tau \\
&= -\frac{1}{i\omega_0}A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}|_0^t \\
&\quad +\frac{1}{i\omega_0}\varepsilon^2\int_0^t \partial_T A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}d\tau \\
&\quad -\frac{1}{(i\omega_0)^2}\varepsilon\partial_X A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}|_0^t \\
&\quad +\frac{1}{(i\omega_0)^2}\varepsilon^3\int_0^t \partial_X\partial_T A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}d\tau \\
&\quad +\frac{1}{(i\omega_0)^2}\varepsilon^2\int_0^t \partial_X^2 A^1(\varepsilon(\alpha - c_g\tau), \varepsilon^2\tau)e^{i(k_0\alpha-\omega_0\tau)}d\tau.
\end{aligned}$$

The integrals involving the other A^js can be treated analogously. Hence, we get

$$\|\varepsilon\int_0^t A^j(\varepsilon(\cdot - c_g\tau), \varepsilon^2\tau)\mathbf{E}^j d\tau\|_{H^s} \leq C(\sqrt{\varepsilon} + t\varepsilon^{5/2})\ .$$

(The "loss" of half a power of ε is again just a reflection of the way in which the Sobolev norms scale when evaluated on long-wavelength functions.)

Combining this with the estimates above, we see that for $0 \leq t \leq T_1/\varepsilon^2$, we have $\|X_1(\cdot,t) - X_1(\cdot,0)\|_{C_b^{s-2}} \leq C\sqrt{\varepsilon}$. Thus, by the inverse function theorem the function $\tilde{X}_1(\alpha,t) = \alpha + X_1(\alpha,t)$ has an inverse $\tilde{X}_1^{-1}(x,t) = x + \Xi(x,t)$ with

$$\sup_{t\in[0,T_1/\varepsilon^2]} \|\Xi(\cdot,t)\|_{C_b^{s-2}} \leq C\sqrt{\varepsilon}.$$

Thus, if we note that $\varepsilon(\psi_1 + \psi_{-1})$ is equal to the order ε term in $\varepsilon\Psi_{X_2}$ we have

$$\begin{aligned}
&\sup_{t\in[0,T_1/\varepsilon^2]} \|\eta(\cdot,t) - \varepsilon(\psi_1(\cdot,t) + \psi_{-1}(\cdot,t))\|_{C_b^{s-2}} \\
&\leq \sup_{t\in[0,T_1/\varepsilon^2]} (\|X_2(\cdot,t) - \varepsilon(\psi_1(\cdot,t) + \psi_{-1}(\cdot,t))\|_{C_b^{s-2}} + \|X_2(\cdot,t) - \eta(\cdot,t)\|_{C_b^{s-2}}) \\
&= \sup_{t\in[0,T_1/\varepsilon^2]} (\|X_2(\cdot,t) - \varepsilon(\psi_1(\cdot,t) + \psi_{-1}(\cdot,t))\|_{C_b^{s-2}} + \|X_2(\cdot,t) - X_2(\tilde{X}_1^{-1}(\cdot,t),t)\|_{C_b^{s-2}}) \\
&\leq C\varepsilon^{3/2} + C\varepsilon^{3/2}\ .
\end{aligned}$$

The estimate on w is similar and Theorem 1.1 follows. □

The rest of this paper is devoted to the proof of Theorem 2.8. The proof consists of an estimate showing that the error function R stays $\mathcal{O}(1)$ bounded on the long time interval of length $\mathcal{O}(1/\varepsilon^2)$. In order to do so, in the equations for the error the terms of $\mathcal{O}(\varepsilon)$ have to be eliminated by a normal-form transform.

3 The normal-form transform

As described in the introduction, in order to control the solutions of the equation for the very long time intervals needed to justify the NLS approximation we must make normal-form transforms to simplify the equations of motion. There is a resonance at the wave number $k = 0$ in this problem which, in a sense that we explain below, is "trivial". However, this resonance implies the existence of a "nontrivial" resonance at the wave number $k = k_0$ which we treat using a slight correction of the method in [Schn98a]. An additional complication in the present situation is due to the "artificial" variable Z_1 which results in one component of the diagonalized system of equations (16) having a linear frequency that is identically zero. As a result we get additional resonances which necessitate rescaling the correction to the NLS equation differently for different wave numbers. This in turn leads to further complications, but in the end, we obtain, as described in the outline of the method, a normal-form transform which results in a linear system whose evolution remains bounded over the time scale of interest.

3.1 The ansatz for the error function

We consider first the diagonalized system

$$
\begin{aligned}
\partial_t c_1 &= B_1(c_1,c_2) + B_2(c_1,c_3) + B_3(c_2,c_3) \\
&\quad + B_4(c_2,c_2) + B_5(c_3,c_3) + \mathcal{O}(\|c\|^3), \\
\partial_t c_2 &= -i\omega c_2 + B_6(c_1,c_2) + B_7(c_1,c_3) + B_8(c_2,c_3) \\
&\quad + B_9(c_2,c_2) + B_{10}(c_3,c_3) + \mathcal{O}(\|c\|^3), \\
\partial_t c_3 &= i\omega c_3 + B_{11}(c_1,c_2) + B_{12}(c_1,c_3) + B_{13}(c_2,c_3) \\
&\quad + B_{14}(c_2,c_2) + B_{15}(c_3,c_3) + \mathcal{O}(\|c\|^3),
\end{aligned}
\tag{40}
$$

associated to (13), where here and in the following the B_j stand for bilinear mappings which do not depend explicitly on α. Notice that we do not have $B(c_1,c_1)$-terms, cf. (16).

The explicit form of these bilinear terms can be computed with the aid of the expansion of the operator $K(X)$ found in [Cr85] and [SW00] and they are listed in Appendix A.3. In Sections 2.3 and 2.4, we computed the formal approximation to the solutions of this system of equations and found:

$$
\begin{aligned}
c_1 &= \varepsilon^2\psi_{q1}, \\
c_2 &= \varepsilon\psi_1 + \varepsilon\psi_{-1} + \varepsilon^2\psi_{q2}, \\
c_3 &= \varepsilon^2\psi_{q3}.
\end{aligned}
$$

If we now write the true solution as the sum of this approximation, plus a correction term, i.e., if we write $c_1 = \varepsilon^2\psi_{q1} + \varepsilon^\beta R_1$, $c_2 = \varepsilon\psi_1 + \varepsilon\psi_{-1} + \varepsilon^2\psi_{q2} + \varepsilon^\beta R_2$ and $c_3 = \varepsilon^2\psi_{q3} + \varepsilon^\beta R_3$, for a $\beta > 1$ sufficiently large, then we find that the equations of motion for the R_j's contain not only the diagonal terms 0, $-i\omega$ and $i\omega$ but also terms linear in R_j and of $\mathcal{O}(\varepsilon)$ of the form $\varepsilon B_1(\psi_1,R_1)$, $\varepsilon B_1(\psi_{-1},R_1)$, etc. Our basic goal is to remove these terms by making normal-form transformations of the form $w_j = R_j + \varepsilon N_j^+(\psi_1,R) + \varepsilon N_j^-(\psi_{-1},R)$ and choosing $N_j^\pm$ to eliminate the terms of $\mathcal{O}(\varepsilon)$ in the equations for R_j. Unfortunately, certain terms are impossible to eliminate by this procedure. For instance, consider the term $B_9(\psi_1,R_2)+B_9(R_2,\psi_1)$. If we write the Fourier transform of this term as $\int \hat{\mathcal{B}}_9(k,k-\ell,\ell)\hat\psi_1(k-\ell)\hat R_2(\ell)d\ell$, and if we write the (Fourier transform of) the corresponding term in the normal-form transformation

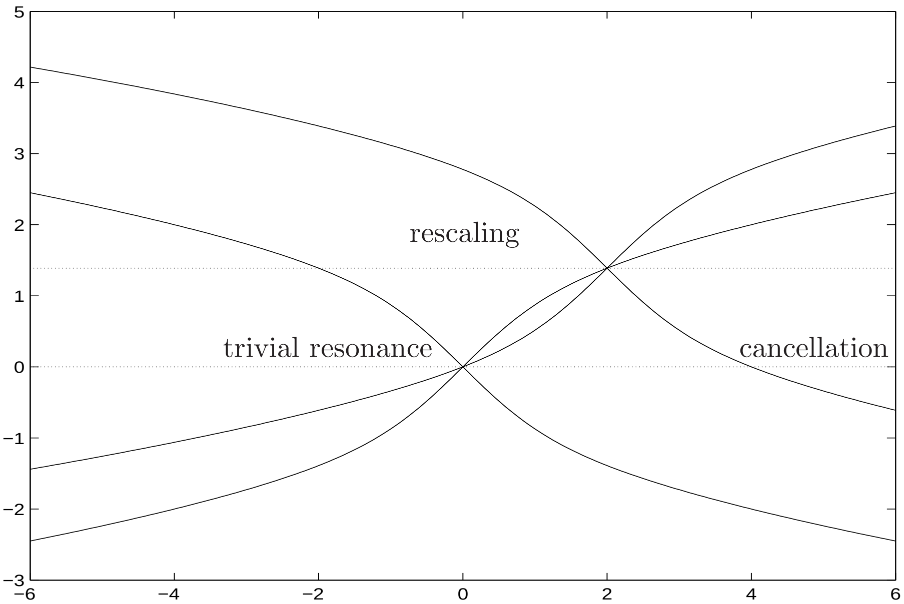

Figure 2: The curves $k \mapsto \omega_j(k)$ and the curves $k \mapsto \omega_3(k_0)+\omega_m(k-k_0)$ for $j, m \in \{1,2,3\}$ and $k_0 = 2$. Intersection points correspond to **quadratic resonances**. There are the resonances at $k=0$ and $k = \pm k_0$ already mentioned in the introduction. The resonance at $k=0$ turns out to trivial, i.e., also the nonlinear terms vanish at this wavenumber. The resonances at $k = \pm k_0$ can only be resolved by scaling c_1, c_2 and c_3 one order smaller close to $k=0$. Beside these resonances there are additional resonances at $\pm 2k_0$ coming from the artificial variable c_1 in the Lagrangian formulation. In the resonances at $\pm 2k_0$ there is some cancellation of terms.

as $\int \hat{N}_9^+(k, k-\ell, \ell)\hat{\psi}_1(k-\ell)\hat{w}(\ell)d\ell$, we find that the kernel in the normal-form transformation must satisfy

$$-i(\omega(k) - \omega(k_0) - \omega(k-k_0))\hat{N}_9^+(k, k-\ell, \ell) = \hat{\mathcal{B}}_9(k, k-\ell, \ell) \ .$$

Unfortunately, the resonance at $k = k_0$ prevents us from solving this equation (or at least, if we solve this equation the expression for $\hat{N}_9^+$ will have a zero in the denominator.) Note that this problem arises from the behavior of R_2 near wave number zero and to circumvent this problem we rescale the error function by an additional power of ε for wave numbers close to zero. Postponing the details of the estimates until later, the problem is solved by making the final ansatz

$$\begin{aligned} c_1 &= \varepsilon^2\psi_{q1} + \varepsilon^3\vartheta R_1, \\ c_2 &= \varepsilon\psi_1 + \varepsilon\psi_{-1} + \varepsilon^2\psi_{q2} + \varepsilon^3\vartheta R_2, \\ c_3 &= \varepsilon^2\psi_{q3} + \varepsilon^3\vartheta R_3, \end{aligned} \tag{41}$$

where ϑR_j is defined by $\widehat{\vartheta R_j} = \hat{\vartheta}\hat{R}_j$ with

$$\hat{\vartheta}(k) = \begin{cases} 1 & \text{for} \quad |k| > \delta\,, \\ \varepsilon + (1-\varepsilon)|k|/\delta & \text{for} \quad |k| \leq \delta \end{cases}$$

with δ chosen as above. By this choice $\hat{\vartheta}(k)\hat{R}_j(k)$ is small at the wavenumbers close to zero reflecting the fact that the nonlinearity vanishes at $k=0$. Moreover, we define $R = (R_1, R_2, R_3)$.

If we ignore the inhomogeneous and nonlinear terms which we have already explained how to handle, we obtain the following system of equations:

$$
\begin{aligned}
\partial_t R_1 &= \vartheta^{-1}\big(\varepsilon B_{1,1,1}(\psi_1,\vartheta R_1) + \varepsilon B_{1,-1,1}(\psi_{-1},\vartheta R_1) + \varepsilon B_{1,1,2}(\psi_1,\vartheta R_2) \\
&\quad +\varepsilon B_{1,-1,2}(\psi_{-1},\vartheta R_2) + \varepsilon B_{1,1,3}(\psi_1,\vartheta R_3) + \varepsilon B_{1,-1,3}(\psi_{-1},\vartheta R_3) \\
&\quad +\varepsilon^2 B_{1,2,1}(\Psi_2,\vartheta R_1) + \varepsilon^2 B_{1,2,2}(\Psi_2,\vartheta R_2) + \varepsilon^2 B_{1,2,3}(\Psi_2,\vartheta R_3) \\
&\quad +\varepsilon^2 B_{1,-2,1}(\Psi_{-2},\vartheta R_1) + \varepsilon^2 B_{1,-2,2}(\Psi_{-2},\vartheta R_2) + \varepsilon^2 B_{1,-2,3}(\Psi_{-2},\vartheta R_3) \\
&\quad +\varepsilon^2 B_{1,0,1}(\Psi_0,\vartheta R_1) + \varepsilon^2 B_{1,0,2}(\Psi_0,\vartheta R_2) + \varepsilon^2 B_{1,0,3}(\Psi_0,\vartheta R_3) \\
&\quad +\varepsilon^2 \mathcal{T}_{1,1,1,1}(\psi_1,\psi_1,\vartheta R_1) + \varepsilon^2 \mathcal{T}_{1,1,1,2}(\psi_1,\psi_1,\vartheta R_2) + \varepsilon^2 \mathcal{T}_{1,1,1,3}(\psi_1,\psi_1,\vartheta R_3) \\
&\quad +\varepsilon^2 \mathcal{T}_{1,-1,-1,1}(\psi_{-1},\psi_{-1},\vartheta R_1) + \varepsilon^2 \mathcal{T}_{1,-1,-1,2}(\psi_{-1},\psi_{-1},\vartheta R_2) \\
&\quad +\varepsilon^2 \mathcal{T}_{1,-1,-1,3}(\psi_{-1},\psi_{-1},\vartheta R_3) + \varepsilon^2 \mathcal{T}_{1,1,-1,1}(\psi_1,\psi_{-1},\vartheta R_1) \\
&\quad +\varepsilon^2 \mathcal{T}_{1,1,-1,2}(\psi_1,\psi_{-1},\vartheta R_2) + \varepsilon^2 \mathcal{T}_{1,1,-1,3}(\psi_1,\psi_{-1},\vartheta R_3) + \mathcal{O}(\varepsilon^3)\big), \\
\partial_t R_2 &= -i\omega R_2 + \vartheta^{-1}\big(\varepsilon B_{2,1,1}(\psi_1,\vartheta R_1) + \varepsilon B_{2,-1,1}(\psi_{-1},\vartheta R_1) + \varepsilon B_{2,1,2}(\psi_1,\vartheta R_2) \\
&\quad +\varepsilon B_{2,-1,2}(\psi_{-1},\vartheta R_2) + \varepsilon B_{2,1,3}(\psi_1,\vartheta R_3) + \varepsilon B_{2,-1,3}(\psi_{-1},\vartheta R_3) \\
&\quad +\varepsilon^2 B_{2,2,1}(\Psi_2,\vartheta R_1) + \varepsilon^2 B_{2,2,2}(\Psi_2,\vartheta R_2) + \varepsilon^2 B_{2,2,3}(\Psi_2,\vartheta R_3) \\
&\quad +\varepsilon^2 B_{2,-2,1}(\Psi_{-2},\vartheta R_1) + \varepsilon^2 B_{2,-2,2}(\Psi_{-2},\vartheta R_2) + \varepsilon^2 B_{2,-2,3}(\Psi_{-2},\vartheta R_3) \\
&\quad +\varepsilon^2 B_{2,0,1}(\Psi_0,\vartheta R_1) + \varepsilon^2 B_{2,0,2}(\Psi_0,\vartheta R_2) + \varepsilon^2 B_{2,0,3}(\Psi_0,\vartheta R_3) \qquad (42)\\
&\quad +\varepsilon^2 \mathcal{T}_{2,1,1,1}(\psi_1,\psi_1,\vartheta R_1) + \varepsilon^2 \mathcal{T}_{2,1,1,2}(\psi_1,\psi_1,\vartheta R_2) + \varepsilon^2 \mathcal{T}_{2,1,1,3}(\psi_1,\psi_1,\vartheta R_3) \\
&\quad +\varepsilon^2 \mathcal{T}_{2,-1,-1,1}(\psi_{-1},\psi_{-1},\vartheta R_1) + \varepsilon^2 \mathcal{T}_{2,-1,-1,2}(\psi_{-1},\psi_{-1},\vartheta R_2) \\
&\quad +\varepsilon^2 \mathcal{T}_{2,-1,-1,3}(\psi_{-1},\psi_{-1},\vartheta R_3) + \varepsilon^2 \mathcal{T}_{2,1,-1,1}(\psi_1,\psi_{-1},\vartheta R_1) \\
&\quad +\varepsilon^2 \mathcal{T}_{2,1,-1,2}(\psi_1,\psi_{-1},\vartheta R_2) + \varepsilon^2 \mathcal{T}_{2,1,-1,3}(\psi_1,\psi_{-1},\vartheta R_3) + \mathcal{O}(\varepsilon^3)\big), \\
\partial_t R_3 &= i\omega R_3 + \vartheta^{-1}\big(\varepsilon B_{3,1,1}(\psi_1,\vartheta R_1) + \varepsilon B_{3,-1,1}(\psi_{-1},\vartheta R_1) + \varepsilon B_{3,1,2}(\psi_1,\vartheta R_2) \\
&\quad +\varepsilon B_{3,-1,2}(\psi_{-1},\vartheta R_2) + \varepsilon B_{3,1,3}(\psi_1,\vartheta R_3) + \varepsilon B_{3,-1,3}(\psi_{-1},\vartheta R_3) \\
&\quad +\varepsilon^2 B_{3,2,1}(\Psi_2,\vartheta R_1) + \varepsilon^2 B_{3,2,2}(\Psi_2,\vartheta R_2) + \varepsilon^2 B_{3,2,3}(\Psi_2,\vartheta R_3) \\
&\quad +\varepsilon^2 B_{3,-2,1}(\Psi_{-2},\vartheta R_1) + \varepsilon^2 B_{3,-2,2}(\Psi_{-2},\vartheta R_2) + \varepsilon^2 B_{3,-2,3}(\Psi_{-2},\vartheta R_3) \\
&\quad +\varepsilon^2 B_{3,0,1}(\Psi_0,\vartheta R_1) + \varepsilon^2 B_{3,0,2}(\Psi_0,\vartheta R_2) + \varepsilon^2 B_{3,0,3}(\Psi_0,\vartheta R_3) \\
&\quad +\varepsilon^2 \mathcal{T}_{3,1,1,1}(\psi_1,\psi_1,\vartheta R_1) + \varepsilon^2 \mathcal{T}_{3,1,1,2}(\psi_1,\psi_1,\vartheta R_2) + \varepsilon^2 \mathcal{T}_{3,1,1,3}(\psi_1,\psi_1,\vartheta R_3) \\
&\quad +\varepsilon^2 \mathcal{T}_{3,-1,-1,1}(\psi_{-1},\psi_{-1},\vartheta R_1) + \varepsilon^2 \mathcal{T}_{3,-1,-1,2}(\psi_{-1},\psi_{-1},\vartheta R_2) \\
&\quad +\varepsilon^2 \mathcal{T}_{3,-1,-1,3}(\psi_{-1},\psi_{-1},\vartheta R_3) + \varepsilon^2 \mathcal{T}_{3,1,-1,1}(\psi_1,\psi_{-1},\vartheta R_1) \\
&\quad +\varepsilon^2 \mathcal{T}_{3,1,-1,2}(\psi_1,\psi_{-1},\vartheta R_2) + \varepsilon^2 \mathcal{T}_{3,1,-1,3}(\psi_1,\psi_{-1},\vartheta R_3) + \mathcal{O}(\varepsilon^3)\big),
\end{aligned}
$$

where

$$
\begin{aligned}
B_{1,\pm1,1}(\psi_{\pm1},\vartheta R_1) &= B_1(\vartheta R_1,\psi_{\pm1})\,,\\
B_{1,\pm1,2}(\psi_{\pm1},\vartheta R_2) &= B_4(\psi_{\pm1},\vartheta R_2)+B_4(\vartheta R_2,\psi_{\pm1})\,,\\
B_{1,\pm1,3}(\psi_{\pm1},\vartheta R_3) &= B_3(\psi_{\pm1},\vartheta R_3)\,,\\
B_{1,\pm2,1}(\Psi_{\pm2},\vartheta R_1) &= B_1(\vartheta R_1,\psi_{(\pm2)2})+B_2(\vartheta R_1,\psi_{(\pm2)3})\,,\\
B_{1,\pm2,2}(\Psi_{\pm2},\vartheta R_2) &= B_1(\psi_{(\pm2)1},\vartheta R_2)+B_3(\vartheta R_2,\psi_{(\pm2)3})+B_4(\psi_{(\pm2)2},\vartheta R_2)\\
&\quad +B_4(\vartheta R_2,\psi_{(\pm2)2})\,,\\
B_{1,\pm2,3}(\Psi_{\pm2},\vartheta R_3) &= B_2(\psi_{(\pm2)1},\vartheta R_3)+B_3(\psi_{(\pm2)2},\vartheta R_3)+B_5(\psi_{(\pm2)3},\vartheta R_3)\\
&\quad +B_5(\vartheta R_3,\psi_{(\pm2)3})\,,\\
B_{1,0,1}(\Psi_0,\vartheta R_1) &= B_1(\vartheta R_1,\psi_{02})+B_2(\vartheta R_1,\psi_{03})\,,\\
B_{1,0,2}(\Psi_0,\vartheta R_2) &= B_1(\psi_{01},\vartheta R_2)+B_3(\vartheta R_2,\psi_{03})+B_4(\psi_{02},\vartheta R_2)+B_4(\vartheta R_2,\psi_{02})\,,\\
B_{1,0,3}(\Psi_0,\vartheta R_3) &= B_2(\psi_{01},\vartheta R_3)+B_3(\psi_{02},\vartheta R_3)+B_5(\psi_{03},\vartheta R_3)+B_5(\vartheta R_3,\psi_{03})\,,\\
B_{2,\pm1,1}(\psi_{\pm1},\vartheta R_1) &= B_6(\vartheta R_1,\psi_{\pm1})\,,\\
B_{2,\pm1,2}(\psi_{\pm1},\vartheta R_2) &= B_9(\psi_{\pm1},\vartheta R_2)+B_9(\vartheta R_2,\psi_{\pm1})\,,\\
B_{2,\pm1,3}(\psi_{\pm1},\vartheta R_3) &= B_8(\psi_{\pm1},\vartheta R_3)\,,\\
B_{2,\pm2,1}(\Psi_{\pm2},\vartheta R_1) &= B_6(\vartheta R_1,\psi_{(\pm2)2})+B_7(\vartheta R_1,\psi_{(\pm2)3})\,,\\
B_{2,\pm2,2}(\Psi_{\pm2},\vartheta R_2) &= B_6(\psi_{(\pm2)1},\vartheta R_2)+B_8(\vartheta R_2,\psi_{(\pm2)3})+B_9(\psi_{(\pm2)2},\vartheta R_2)\\
&\quad +B_9(\vartheta R_2,\psi_{(\pm2)2})\,,\\
B_{2,\pm2,3}(\Psi_{\pm2},\vartheta R_3) &= B_7(\psi_{(\pm2)1},\vartheta R_3)+B_8(\psi_{(\pm2)2},\vartheta R_3)+B_{10}(\psi_{(\pm2)3},\vartheta R_3)\\
&\quad +B_{10}(\vartheta R_3,\psi_{(\pm2)3})\,,\\
B_{2,0,1}(\Psi_0,\vartheta R_1) &= B_6(\vartheta R_1,\psi_{02})+B_7(\vartheta R_1,\psi_{03})\,,\\
B_{2,0,2}(\Psi_0,\vartheta R_2) &= B_6(\psi_{01},\vartheta R_2)+B_8(\vartheta R_2,\psi_{03})+B_9(\psi_{02},\vartheta R_2)+B_9(\vartheta R_2,\psi_{02})\,,\\
B_{2,0,3}(\Psi_0,\vartheta R_3) &= B_7(\psi_{01},\vartheta R_3)+B_8(\psi_{02},\vartheta R_3)+B_{10}(\psi_{03},\vartheta R_3)+B_{10}(\vartheta R_3,\psi_{03})\,,\\
B_{3,\pm1,1}(\psi_{\pm1},\vartheta R_1) &= B_{11}(\vartheta R_1,\psi_{\pm1})\,,\\
B_{3,\pm1,2}(\psi_{\pm1},\vartheta R_2) &= B_{14}(\psi_{\pm1},\vartheta R_2)+B_{14}(\vartheta R_2,\psi_{\pm1})\,,\\
B_{3,\pm1,3}(\psi_{\pm1},\vartheta R_3) &= B_{13}(\psi_{\pm1},\vartheta R_3)\,,\\
B_{3,\pm2,1}(\Psi_{\pm2},\vartheta R_1) &= B_{11}(\vartheta R_1,\psi_{(\pm2)2})+B_{12}(\vartheta R_1,\psi_{(\pm2)3})\,,\\
B_{3,\pm2,2}(\Psi_{\pm2},\vartheta R_2) &= B_{11}(\psi_{(\pm2)1},\vartheta R_2)+B_{13}(\vartheta R_2,\psi_{(\pm2)3})+B_{14}(\psi_{(\pm2)2},\vartheta R_2)\\
&\quad +B_{14}(\vartheta R_2,\psi_{(\pm2)2})\,,\\
B_{3,\pm2,3}(\Psi_{\pm2},\vartheta R_3) &= B_{12}(\psi_{(\pm2)1},\vartheta R_3)+B_{13}(\psi_{(\pm2)2},\vartheta R_3)+B_{15}(\psi_{(\pm2)3},\vartheta R_3)\\
&\quad +B_{15}(\vartheta R_3,\psi_{(\pm2)3})\,,\\
B_{3,0,1}(\Psi_0,\vartheta R_1) &= B_{11}(\vartheta R_1,\psi_{02})+B_{12}(\vartheta R_1,\psi_{03})\,,\\
B_{3,0,2}(\Psi_0,\vartheta R_2) &= B_{11}(\psi_{01},\vartheta R_2)+B_{13}(\vartheta R_2,\psi_{03})+B_{14}(\psi_{02},\vartheta R_2)+B_{14}(\vartheta R_2,\psi_{02})\,,\\
B_{3,0,3}(\Psi_0,\vartheta R_3) &= B_{12}(\psi_{01},\vartheta R_3)+B_{13}(\psi_{02},\vartheta R_3)+B_{15}(\psi_{03},\vartheta R_3)+B_{15}(\vartheta R_3,\psi_{03})
\end{aligned}
$$

and the terms $\mathcal{T}_{i,j,k,l}(\psi_j,\psi_k,\vartheta R_l)$ stand for trilinear terms which are linear in the R_l. They will be treated in detail later. The operator ϑ^{-1} is defined by its symbol $\widehat{\vartheta^{-1}}(k)=\hat{\vartheta}^{-1}(k)=(\hat{\vartheta}(k))^{-1}$. Notice that $\hat{\vartheta}^{-1}(k)$ is at most of order $\mathcal{O}(\varepsilon^{-1})$ for $|k|\le\delta$ but of order $\mathcal{O}(1)$ for $|k|\ge\delta$. We note here that the arguments of the terms $B_{j,0,n}$ and $B_{j,2,n}$ are written as Ψ rather than ψ to emphasize that these terms depend on more than one component of the vector valued approximating function Ψ, whereas the remaining terms depend only on a single, scalar component, and hence their arguments are denoted ψ_j.

3.2 The normal-form strategy for an example

In order to eliminate the $\mathcal{O}(1)$ and $\mathcal{O}(\varepsilon)$ terms on the right hand sides of these equations we make a series of normal-form transformations. Each of these transformations is a near identity transformation which removes one (or more) of the "bad" terms.

We will explain the general strategy with a quite specific example which illustrates the general principles involved and then go through each of the many terms that must be removed from the right hand side of these equations and explain in turn how they are treated.

Suppose, for example, that we consider the terms from the second equation in (42)

$$\partial_t R_2 = -i\omega R_2 + \varepsilon\vartheta^{-1} B_{2,1,2}(\psi_1, \vartheta R_2) + \dots \tag{43}$$

and we wish to eliminate the term $\varepsilon\vartheta^{-1}B_{2,1,2}(\psi_1,\vartheta R_2) = \varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2) + B_9(\vartheta R_2,\psi_1))$ from the equation.

We will write $\tilde{R}_2 = R_2 + \varepsilon N(\psi_1, R_2)$ with N a bilinear function chosen in such a way that the term $\varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2)+B_9(\vartheta R_2,\psi_1))$ does not appear in the evolution equation for $\tilde{R}_2$. If we write the evolution equation for $\tilde{R}_2$, we find

$$\begin{aligned}
\partial_t \tilde{R}_2 &= \partial_t R_2 + \varepsilon N(\partial_t\psi_1, R_2) + \varepsilon N(\psi_1, \partial_t R_2) \\
&= -i\omega R_2 + \varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2) + B_9(\vartheta R_2,\psi_1)) + \varepsilon N(\partial_t\psi_1, R_2) \\
&\quad + \varepsilon N(\psi_1, \partial_t R_2) + \dots \\
&= -i\omega \tilde{R}_2 + \varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2) + B_9(\vartheta R_2,\psi_1)) + i\omega\varepsilon N(\psi_1, R_2) \\
&\quad + \varepsilon N(\partial_t\psi_1, R_2) + \varepsilon N(\psi_1, \partial_t R_2) + \dots
\end{aligned} \tag{44}$$

Thus, we see that in order to eliminate the term $\varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2)+B_9(\vartheta R_2,\psi_1))$ we should choose the normal-form transformation N to satisfy

$$\varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2)+B_9(\vartheta R_2,\psi_1))+i\omega\varepsilon N(\psi_1,R_2)+\varepsilon N(\partial_t\psi_1,R_2)+\varepsilon N(\psi_1,\partial_t R_2)=0\ . \tag{45}$$

We can simplify this slightly if we replace $\partial_t R_2$ in the argument of the last term by $-i\omega R_2$, which we can do at the expense of introducing additional terms of $\mathcal{O}(\varepsilon^2)$, which are absorbed in the terms we have already neglected in the equation for $\partial_t R_2$. We would like to make a similar replacement of $\partial_t\psi_1$ but if we compute this derivative using the explicit expression for ψ_1 we find it is not quite equal to $-i\omega\psi_1$. However, as we will show in Lemma 3.3 below, it can be approximated by this expression – i.e., $\partial_t\psi_1 = -i\omega\psi_1 + \mathcal{O}(\varepsilon^2)$. If we make this substitution (and again absorb the error in the terms we have ignored we find that N should satisfy:

$$\varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2)+B_9(\vartheta R_2,\psi_1))+i\omega\varepsilon N(\psi_1,R_2)-\varepsilon N(i\omega\psi_1,R_2)-\varepsilon N(\psi_1,i\omega R_2)=0\ . \tag{46}$$

From the explicit formulas for the quadratic terms in (40) we see that we can write the Fourier transform of the bilinear term $\vartheta^{-1}(B_9(\psi_1,\vartheta R_2) + B_9(\vartheta R_2,\psi_1))$ as

$$\int \hat{\vartheta}^{-1}(k)\hat{\mathcal{B}}_9(k,k-\ell,\ell)\hat{\psi}_1(k-\ell)\hat{\vartheta}(\ell)\hat{R}_2(\ell)d\ell$$

for some appropriate kernel $\hat{\mathcal{B}}_9$ whose explicit form we discuss below. In order to compute the kernel $\hat{n}$ of N, we write out the Fourier transform of (46) which gives

$$\begin{aligned} & -\int \hat{\vartheta}^{-1}(k)\hat{\mathcal{B}}_9(k,k-\ell,\ell)\hat{\vartheta}(\ell)\hat{\psi}_1(k-\ell)\hat{R}_2(\ell)d\ell \\ = \quad & i\omega(k)\int \hat{n}(k,k-\ell,\ell)\hat{\psi}_1(k-\ell)\hat{R}_2(\ell)d\ell \\ & -i\int \hat{n}(k,k-\ell,\ell)\omega(k-\ell)\hat{\psi}_1(k-\ell)\hat{R}_2(\ell)d\ell \\ & -i\int \hat{n}(k,k-\ell,\ell)\hat{\psi}_1(k-\ell)\omega(\ell)\hat{R}_2(\ell)d\ell \ , \end{aligned} \tag{47}$$

or, eliminating the integrals and focusing on the equation satisfied by the kernels we see that $\hat{n}$ should satisfy

$$\hat{n}(k,k-\ell,\ell) \quad = \quad \frac{i\hat{\mathcal{B}}_9(k,k-\ell,\ell)}{(\omega(k)-\omega(k-\ell)-\omega(\ell))}\frac{\hat{\vartheta}(\ell)}{\hat{\vartheta}(k)} . \tag{48}$$

Clearly, $\hat{n}$ will only be defined if the expression $(\omega(k)-\omega(k-\ell)-\omega(\ell))$ can be bounded away from zero (or in some rare cases which we discuss below, if a zero in this expression is off-set by a zero of $\hat{\mathcal{B}}_9$ at the same values of k and ℓ.) This requirement leads to our non-resonance conditions which we will have to verify in each of many different possible cases below. Before beginning this straightforward but lengthy procedure we make a few more general remarks:

1. We must eliminate not only the term $\varepsilon\vartheta^{-1}(B_9(\psi_1,\vartheta R_2)+B_9(\vartheta R_2,\psi_1))$ from the equation for $\partial_t R_2$, but also many other terms – and even more in the equations for the other variables. However, since the terms to be eliminated are linear in the dependent variables R_j, the transformations will be linear as well and we can construct the final transformation in lowest order as a sum of the transformations constructed to eliminate each of the terms in turn.

2. In order to prove our approximation theorems we will have to show not only that the kernel $\hat{n}$ is well defined but also to estimate how the transformation it defines acts on the function spaces $Y_{\sigma,s}$. This we will do with the aid of Lemma 3.1 below.

3. We must also eliminate a few terms of the form $\vartheta^{-1}B_n(\partial_t\psi_{qk},\vartheta R_j)$. These must be handled in a slightly different fashion because $\partial_t\hat{\psi}_{qk}$ is not approximated by $i\omega_k\hat{\psi}_{qk}$. We explain exactly what modifications are necessary in our general scheme when we encounter these terms below.

3.3 Some technical lemmas

A key lemma in what follows will be the following estimate of expressions like

$$\hat{N}(\psi_j,R)(k) = \int \hat{n}(k,k-\ell,\ell)\hat{\psi}_j(k-\ell)\hat{R}(\ell)d\ell \ , \tag{49}$$

where ψ_j is a part of the approximation constructed above (and in particular, is an entire function) and R is an element of the space $Y_{\sigma,s}$.

Lemma 3.1. *If there exist constants $C_N > 0$, $s_1, s_2 \geq 0$ and such that*

$$(1+k^2)^{s/2} |\hat{n}(k, k-\ell, \ell)| \leq C_N (1 + |k-\ell|)^{s_1} (1 + |\ell|)^{s_2}$$

Then the expression (49) defines a bounded linear transformation

$$\begin{aligned} N : Y_{\sigma,s_2} &\to Y_{\sigma,s} \\ w &\to N(\psi_j, w) \ , \end{aligned}$$

and there exists a constant C_ψ depending on σ, s, s_1, s_2 and the norm of ψ_j (but independent of ε) such that

$$\|N(\psi_j, w)\|_{Y_{\sigma,s}} \leq C_\psi \|w\|_{Y_{\sigma,s_2}} \ .$$

Proof. From the definition of the norm on $Y_{\sigma,s}$, we have:

$$\begin{aligned} \|N(\psi_j, w)\|^2_{Y_{\sigma,s}} &= \int (1+|k|^2)^s e^{2\sigma|k|} \left(\int \hat{n}(k, k-\ell, \ell) \hat{\psi}_j(k-\ell) \hat{w}(\ell) d\ell \right)^2 dk \\ &\leq C \int \left(\int (1+k^2)^{s/2} e^{\sigma|k|} |\hat{n}(k, k-\ell, \ell)| |\hat{\psi}_j(k-\ell)| |\hat{w}(\ell)| d\ell \right)^2 dk \\ &\leq C C_N \int \left(\int (1+|k-\ell|^2)^{s_1/2} e^{\sigma|k-\ell|} |\hat{\psi}_j(k-\ell)| (1+|\ell|^2)^{s_2/2} e^{\sigma|\ell|} |\hat{w}(\ell)| d\ell \right)^2 dk \ . \end{aligned}$$

If we now think of the k-integration as the square of the L^2-norm of a convolution we can bound it with the aid of Young's inequality by

$$C \|w\|_{Y_{\sigma,s_2}} \int (1+|k|^2)^{s_1/2} e^{\sigma|k|} |\hat{\psi}_j(k)| dk \ .$$

But since ψ_j is entire this last integral is bounded by a constant, and using the way ψ_j is constructed, we see that this constant is independent of ε. $\square$

We will also need to construct normal-form transformations that eliminate trilinear terms in some of these equations. To bound the resulting transformations we use the following lemma whose proof we leave as an exercise since it is a very easy modification of the preceding one.

Lemma 3.2. *Let ψ be entire and $s > 0$ be fixed. Suppose that there exist constants C, and s_j, $j = 1, 2, 3$, such that*

$$(1+k^2)^{s/2} |\hat{M}(k, k-\ell, \ell-p, p)| \leq C (1+|k-\ell|^2)^{s_1/2} (1+|\ell-p|^2)^{s_2/2} (1+p^2)^{s_3/2} \ .$$

Then the mapping $M : w \to M(\psi, \phi, w)$ defined by the kernel $\hat{M}$ is a bounded transformation from Y_{σ,s_3} to $Y_{\sigma,s}$.

We next turn to the result mentioned in our overview of the construction of the normal-forms, namely the fact that we can approximate $\partial_t \psi_1$ by $-i\omega\psi_1$.

Lemma 3.3. *Fix $s, \sigma > 0$. There exists a constant $C_\psi = C_\psi(s, \sigma) > 0$ such that*

$$\|\partial_t \psi_{\pm 1} + i\omega \psi_{\pm 1}\|_{Y_{\sigma,s}} \leq C_\psi \varepsilon^2$$

Before proving this lemma we note that if we combine it with the method of proof of Lemma 3.1 we easily obtain:

Corollary 3.4. *If there exist constants C_b, s_1 and s_2 such that*

$$(1+k^2)^{s/2}|\hat{b}(k,k-\ell,\ell)| \leq C_b(1+|k-\ell|)^{s_1}(1+|\ell|)^{s_2} ,$$

then there exists a constant C_B (independent of ε) such that the bilinear term $B(\psi_{\pm 1}, R)$ defined by the kernel $\hat{b}$ satisfies

$$\|B(\partial_t \psi_{\pm 1} + i\omega\psi_{\pm 1}, R)\|_{Y_{\sigma,s}} \leq C_B\, \varepsilon^2 \|R\|_{Y_{\sigma,s_2}} .$$

Proof. (of Lemma 3.3) We will prove the case of ψ_1 – the case of ψ_{-1} works analogously by the change $k_0 \to -k_0$. From the explicit formulas for ψ_1 we have

$$\partial_t \hat{\psi}_1(k,t) + i\omega(k)\hat{\psi}_1(k,t) = i(-\omega(k_0) - c_g(k-k_0) + \omega(k))\hat{\psi}_1(k,t) + \varepsilon^2 \partial_T \hat{\psi}_1(k,t) .$$

Here, the partial derivative with respect to "T" refers to differentiation with respect to the "slow" time $\varepsilon^2 t$ that occurs in the second argument of the amplitude function A_1 in the definition of ψ_1. First consider the norm of the term $\partial_T \hat{\psi}_1$. Recalling (see (35)) that we truncated the support of A_1 in a neighborhood of size δ we see that

$$\begin{aligned} \int (1+|k|^2)^{s_1/2} e^{\sigma|k|} |\partial_T \hat{\psi}_1(k,t)| dk &\leq \frac{1}{\varepsilon} \int_{\frac{|k-k_0|}{\varepsilon} \leq \delta} (1+|k|^2)^{s_1/2} e^{\sigma|k|} |\partial_T \hat{A}_1(\frac{k-k_0}{\varepsilon}, \varepsilon^2 t)| dk \\ &\leq \int_{p=-\delta}^{\delta} (1+|k_0+\varepsilon p|^2)^{s_1/2} e^{\sigma|k_0+\varepsilon p|} |\partial_T \hat{A}_1(p,\varepsilon^2 t)| dp \\ &\leq C(k_0,\sigma) \int_{p=-\delta}^{\delta} (a+2\varepsilon|p|)^{s_1/2} e^{\varepsilon\sigma|p|} |\partial_T \hat{A}_1(p,\varepsilon^2 t)| dp . \end{aligned}$$

However, because of the compact support of $\hat{A}_1$, this integral is finite.

Next note that there exists C_ω such that

$$|\omega(k) - \omega(k_0) - c_g(k-k_0)| \leq C_\omega |k-k_0|^2 .$$

Thus,

$$\begin{aligned} &\int (1+|k|^2)^{s_1/2} e^{\sigma|k|} |\omega(k_0) + c_g(k-k_0) - \omega(k)||\hat{\psi}_1(k,t)| dk \\ &\leq \frac{C_\omega}{\varepsilon} \int |k-k_0|^2 |\hat{A}_1(\frac{k-k_0}{\varepsilon}, \varepsilon^2 t)| dk \\ &\leq C_\omega \varepsilon^2 \int p^2 |\hat{A}_1(p,\varepsilon^2 t)| dp \\ &\leq C(k_0, A_1)\varepsilon^2 , \end{aligned}$$

where in the second inequality we again used the fact that the support of $\hat{\psi}_1$ is bounded and in the last inequality that $\hat{A}_1$ is bounded at least in H^3 . □

With these technical lemmas in hand we now construct our normal-form transformation. The transform is constructed term by term following the outline sketched in equations (44)-(48).

3.4 The first normal-form transform

We begin discussing the normal-form transform for the error equations. Due to the structure of the nonlinear terms in the error equations the size of the Fourier transform of these terms depends on whether k is close to zero or not. In order to separate the behavior in these two regions more clearly we define projection operators P^0 and P^1 by the Fourier multipliers

$$\widehat{P}^0(k) = \chi_{|k|\leq\delta}(k) \quad\text{and}\quad \widehat{P}^1(k) = 1 - \widehat{P}^0(k) \tag{50}$$

for a $\delta > 0$ sufficiently small, but independent of $0 < \varepsilon \ll 1$. (This is the same constant δ that appears in the definition of ϑ.) When necessary we will write

$$R = R^0 + R^1 \ ,$$

with $R^r = P^r R$ for $r = 0, 1$ and analogously with the other variables.

Applying the projection operators $P^{0,1}$ to system (42) we see that it is equivalent to the system of equations

$$\begin{aligned}
\partial_t R_j^0 \;&=\; i\omega_j R_j^0 + \vartheta^{-1} P^0 \Big(\sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon B_{j,l,n}(\psi_l, \vartheta R_n^1) \\
&\quad + \sum_{\substack{l=-2,2\\ n=1,2,3}} \varepsilon^2 B_{j,l,n}(\Psi_l, \vartheta R_n^1) + \sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon^2 \mathcal{T}_{j,l,l,n}(\psi_l, \psi_l, \vartheta R_n^1) \\
&\quad + \sum_{r=0,1} \Big(\sum_{n=1,2,3} \varepsilon^2 B_{j,0,n}(\Psi_0, \vartheta R_n^r) + \sum_{n=1,2,3} \varepsilon^2 \mathcal{T}_{j,1,-1,n}(\psi_1, \psi_{-1}, \vartheta R_n^r) \Big) \\
&\quad + \mathcal{O}(\varepsilon^3) \Big),
\end{aligned}$$

$$\begin{aligned}
\partial_t R_j^1 \;&=\; i\omega_j R_j^1 + \sum_{r=0,1} \vartheta^{-1} P^1 \Big(\sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon B_{j,l,n}(\psi_l, \vartheta R_n^r) \\
&\quad + \sum_{\substack{l=-2,0,2\\ n=1,2,3}} \varepsilon^2 B_{j,l,n}(\Psi_l, \vartheta R_n^r) + \sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon^2 \mathcal{T}_{j,l,l,n}(\psi_l, \psi_l, \vartheta R_n^r) \\
&\quad + \sum_{n=1,2,3} \varepsilon^2 \mathcal{T}_{j,1,-1,n}(\psi_1, \psi_{-1}, \vartheta R_n^r) + \mathcal{O}(\varepsilon^3) \Big)
\end{aligned} \tag{51}$$

for $j = 1, 2, 3$.

Here we used the fact that due to $\widehat{\Psi}_l(k-\ell) = 0$ unless $|(k-\ell) - lk_0| < \delta$ and $\widehat{R}^0(\ell) = 0$ for $|\ell| > \delta$ we have $P^0 B_{j,\pm1,n}(\psi_{\pm1}, \vartheta R_n^0) = P^0 B_{j,\pm2,n}(\Psi_{\pm2}, \vartheta R_n^0) = P^0 \mathcal{T}_{j,\pm1,\pm1,n}(\psi_{\pm1}, \psi_{\pm1}, \vartheta R_n^0) = 0$ for $j, n \in \{1, 2, 3\}$ if $\delta > 0$ is sufficiently small, but independent of $0 < \varepsilon \ll 1$.

Since $\vartheta^{-1} P^1$ is of order $\mathcal{O}(1)$ all terms on the second and the last line of the evolution equation for R_j^1 are at least of order $\mathcal{O}(\varepsilon^2)$ and need not to be eliminated. Moreover, we will show in Subsection 3.7 that all terms on the third line of the evolution equation for R_j^0 are at least of order $\mathcal{O}(\varepsilon^2)$ as well and need not to be eliminated either.

In order to eliminate the terms in (51) which are of order $\mathcal{O}(1)$ or $\mathcal{O}(\varepsilon)$ we look for a normal-form transformation of the form

$$\begin{aligned}
\tilde{R}^0 \;&=\; R^0 + \varepsilon N^{0,1}(\Psi, R^1)\,, \\
\tilde{R}^1 \;&=\; R^1 + \varepsilon N^{1,0}(\Psi, R^0) + \varepsilon N^{1,1}(\Psi, R^1)\,,
\end{aligned}$$

where

$$
\begin{aligned}
\tilde{R}_j^0 &= R_j^0 + \sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon N_{j,l,n}^{0,1}(\psi_l, R_n^1) + \sum_{\substack{l=-2,2\\ n=1,2,3}} \varepsilon^2 N_{j,l,n}^{0,1}(\Psi_l, R_n^1) \\
&\quad + \sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon^2 N_{j,l,l,n}^{0,1}(\psi_l, \psi_l, R_n^1) \ , \\
\tilde{R}_j^1 &= R_j^1 + \sum_{r=0,1} \sum_{\substack{l=-1,1\\ n=1,2,3}} \varepsilon N_{j,l,n}^{1,r}(\psi_l, R_n^r) \ .
\end{aligned} \tag{52}
$$

Construction of $\mathbf{N^{0,1}}$: Now, we start discussing systematically the construction of all components of this normal-form transformation. First, we address the components $N_{j,\pm 1,n}^{0,1}$ for $j, n \in \{1,2,3\}$. Proceeding as in Subsection 3.2 we see that the kernels $\hat{n}_{j,\pm 1,n}^{0,1}$ of $N_{j,\pm 1,n}^{0,1}$ should be of the form

$$
\hat{n}_{j,\pm 1,n}^{0,1}(k, k-\ell, \ell) = \frac{i\widehat{P}^0(k)\hat{b}_{j,\pm 1,n}(k, k-\ell, \ell)}{(-\omega_j(k) - \omega(k-\ell) + \omega_n(\ell))} \frac{\hat{\vartheta}(\ell)}{\hat{\vartheta}(k)} , \tag{53}
$$

where $\hat{b}_{j,\pm 1,n}$ are the kernels of $B_{j,\pm 1,n}$. Due to the fact that the $\widehat{P}^0$ and $\hat{\psi}_{\pm 1}$ have supports localized near $k = 0$ and $(k-\ell) = \pm k_0$ respectively this expression only has to be analyzed for $|(k-\ell) \pm k_0| < \delta$ and $|k| < \delta$. As a consequence for $\delta > 0$ sufficiently small, but independent of $0 < \varepsilon \ll 1$, we can also restrict to wave numbers ℓ bounded away from 0. Hence from the possible resonances discussed above only the resonance at $k = 0$ will play a role for $N_{j,\pm 1,n}^{0,1}$. The kernel $\hat{n}_{j,\pm 1,n}^{0,1}$ can then be estimated as follows. First note that if we consider the denominator of this expression near $k = 0$ then we have

$$
-\omega_j(k) - \omega(k-\ell) + \omega_n(\ell) = -\omega_j'(0)k - (\omega(-\ell) + \omega'(-\ell)k) + \omega_n(\ell) + \mathcal{O}(k^2) \ .
$$

If $\omega_n(\ell) \neq -\omega(\ell)$ this quantity is bounded below by some $\mathcal{O}(1)$ constant for all $|k| < \delta$. If, on the other hand, $\omega_n(\ell) = -\omega(\ell)$, which is true if and only if $n = 2$, there exists a positive constant C such that

$$
| - \omega_j(k) - \omega(k-\ell) + \omega_n(\ell)| \geq C|k| \ . \tag{54}
$$

Here, we have used the fact that $\ell \approx \pm k_0$ because of the support of $\hat{\psi}_{\pm 1}$ and the fact that $\omega'(\pm k_0)$ is $\mathcal{O}(1)$ and is not equal to $\omega_j'(0)$. Thus, only for $n = 2$, does the denominator of the expression for $\hat{n}_{j,\pm 1,n}^{0,1}$ get close to zero. However, in the case $n = 2$ we see from the definitions of the form of the nonlinear terms that $\hat{b}_{j,\pm 1,2}(k, k-\ell, \ell) = \hat{\mathcal{B}}_i(k, k-\ell, \ell)$ for some $i \in \{4, 9, 14\}$. Thus we get by the subsequent Lemma 3.11 that

$$
|\hat{b}_{j,\pm 1,2}(k, k-\ell, \ell)| \leq C|k| \ . \tag{55}
$$

Hence, in the case of $n = 2$, there is a cancellation between the numerator and the denominator, while for other values of n the denominator is bounded away from zero and thus, there exists a constant $C \geq 0$ such that

$$
|\hat{\vartheta}(k)\hat{n}_{j,\pm 1,n}^{0,1}(k, k-\ell, \ell)| \leq C \tag{56}
$$

for all $|k| \leq \delta$ and ℓ under consideration.

Because of the factor of $\widehat{P}^0(k)$ which makes $\hat{n}^{0,1}_{j,\pm1,n}(k,k-\ell,\ell)=0$ if $|k|>\delta$, $N^{0,1}_{j,\pm1,n}$ is "smoothing" in the sense that if $R^1_n\in Y_{\sigma,s}$ for some $s>1$, then given any σ',s', there exists $C_{\sigma',s'}$ such that

$$\|\varepsilon N^{0,1}_{j,\pm1,n}(\psi_{\pm1},R^1_n)\|_{Y_{\sigma',s'}}\le C_{\sigma',s'}\|R^1_n\|_{Y_{\sigma,s}}\ . \tag{57}$$

In particular, this estimate holds when $\sigma'=\sigma$ and $s'=s$. Note, however, that in spite of the factor of ε in front of $N^{0,1}_{j,\pm1,n}$, we cannot assume that $C_{\sigma',s'}\sim\mathcal{O}(\varepsilon)$ because of the factor of $\vartheta^{-1}(k)\sim\varepsilon^{-1}$ for $k\approx0$, in the formula for the kernel of $N^{0,1}_{j,\pm1,n}$.

Now, we address the components $N^{0,1}_{j,\pm2,n}$. Their kernels $\hat{n}^{0,1}_{j,\pm2,n}$ should be of the form

$$\hat{n}^{0,1}_{j,\pm2,n}(k,k-\ell,\ell)\ =\ \frac{i\widehat{P}^0(k)\hat{b}_{j,\pm2,n}(k,k-\ell,\ell)}{(-\omega_j(k)\pm2\omega(k_0)+\omega_n(\ell))}\,\frac{\hat{\vartheta}(\ell)}{\hat{\vartheta}(k)}\,, \tag{58}$$

where we have used the fact that $\partial_t\hat{\Psi}_{\pm2}=\pm2i\omega(k_0)\hat{\Psi}_{\pm2}+\mathcal{O}(\varepsilon)$ to approximate the denominator. Since $k-\ell\approx\pm2k_0$ we can further approximate the denominator as

$$-\omega_j(k)\pm2\omega(k_0)+\omega_n(k\mp2k_0)\ .$$

Because of $\omega(2k_0)\neq\pm2\omega(k_0)\neq0$ the denominator is bounded away from zero for all $|k|\le\delta$. Moreover, since k and ℓ are restricted to bounded intervals, the operators $\varepsilon N^{0,1}_{j,\pm2,n}$ define bounded transformations from $Y_{\sigma,s}$ to itself.

Finally, we address the components $N^{0,1}_{j,\pm1,\pm1,n}$. Proceeding analogously as in the case of the bilinear terms we find that the kernels $\hat{n}^{0,1}_{j,\pm1,\pm1,n}$ should be of the form

$$\hat{n}^{0,1}_{j,\pm1,\pm1,n}(k,k-\ell,\ell-p,p)\ =\ \frac{i\widehat{P}^0(k)\hat{T}_{j,\pm1,\pm1,n}(k,k-\ell,\ell-p,p)}{(-\omega_j(k)-\omega(k-\ell)-\omega(\ell-p)+\omega_n(p))}\,\frac{\hat{\vartheta}(p)}{\hat{\vartheta}(k)}\,, \tag{59}$$

where $\hat{T}_{j,\pm1,\pm1,n}(k,k-\ell,\ell-p,p)$ is the kernel of $T_{j,\pm1,\pm1,n}$. Since $k-\ell\approx\pm k_0$ and $\ell-p\approx\pm k_0$ we can further approximate the denominator as

$$-\omega_j(k)\mp2\omega(k_0)+\omega_n(k\mp2k_0)$$

and therefore the denominator is bounded away from zero for all $|k|\le\delta$. Moreover, since k, ℓ and p are restricted to bounded intervals, the operators $\varepsilon N^{0,1}_{j,\pm1,\pm1,n}$ define bounded transformations from $Y_{\sigma,s}$ to itself.

Construction of $\mathbf{N^{1,0}}$ and $\mathbf{N^{1,1}}$: Before we proceed constructing the normal-form transformation we will replace the terms $\varepsilon\vartheta^{-1}P^1B_{j,\pm1,n}(\psi_{\pm1},\vartheta R^r_n)$ with $j,n\in\{1,2,3\}$ and $r\in\{0,1\}$ in the evolution equations for R^1_j by $\varepsilon\vartheta^{-1}P^1B_{j,\pm1,n}(\psi_{\pm1},\vartheta_0R^r_n)$, where $\hat{\vartheta}_0(k)=\hat{\vartheta}(k)-\varepsilon$. This modification will help us to avoid a resonance problem at $\pm k_0$. The key fact that we will use below is that $\hat{\vartheta}_0(0)=0$. Making this change introduces additional error terms $\varepsilon^2\vartheta^{-1}P^1B_{j,\pm1,n}(\psi_{\pm1},R^r_n)$ into the evolution equations for R^1_j. However, since $\hat{\vartheta}^{-1}(k)$ is $\mathcal{O}(1)$ on the support of $\hat{P}^1$, these terms can be included in the error terms of order $\mathcal{O}(\varepsilon^2)$.

Now, proceeding as in Subsection 3.2 yields that the components $N^{1,r}_{j,\pm1,n}$ should satisfy the equation

$$\begin{aligned}-i\omega_jN^{1,r}_{j,\pm1,n}(\psi_{\pm1},R^r_n)-N^{1,r}_{j,\pm1,n}(i\omega\psi_{\pm1},R^r_n)+N^{1,r}_{j,\pm1,n}(\psi_{\pm1},i\omega_nR^r_n)&\\ =-\vartheta^{-1}P^1B_{j,\pm1,n}(\psi_{\pm1},\vartheta_0R^r_n)\,.&\end{aligned} \tag{60}$$

To extract the real 'dangerous' terms from $\vartheta^{-1}P^1B_{j,\pm1,n}(\psi_{\pm1},\vartheta_0R^r_n)$ we will use the following lemma which takes advantage of the strong localization of $\psi_{\pm1}$ near the wave numbers $\pm k_0$ in Fourier space.

Lemma 3.5. *Fix $p \in I\!R$. Assume that $\kappa \in C(I\!R^3, \mathbb{C})$. Assume further that $\psi \in C^2(I\!R)$ has a finitely supported Fourier transform and that $R \in Y_{\sigma,s}$.*

- *If κ is Lipschitz with respect to its second argument for $k - \ell$ in some neighborhood of $p \in I\!R$, then there exists $C_{\psi,\kappa,p} > 0$ such that*

$$\| \int \kappa(\cdot, \cdot - \ell, \ell)\varepsilon^{-1}\widehat{\psi}(\frac{\cdot - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell - \int \kappa(\cdot, p, \ell)\varepsilon^{-1}\widehat{\psi}(\frac{\cdot - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell\|_{Y_{\sigma,s}} \leq C_{\psi,\kappa,p}\varepsilon\|R\|_{Y_{\sigma,s}} \tag{61}$$

- *If κ is globally Lipschitz with respect to its third argument, then there exists $D_{\psi,\kappa} > 0$ such that*

$$\begin{aligned} &\| \int \kappa(\cdot, \cdot - \ell, \ell)\varepsilon^{-1}\widehat{\psi}(\frac{\cdot - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell \\ &\qquad - \int \kappa(\cdot, \cdot - \ell, \cdot - p)\varepsilon^{-1}\widehat{\psi}(\frac{\cdot - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell\|_{Y_{\sigma,s}} \\ &\leq D_{\psi,\kappa}\varepsilon\|R\|_{Y_{\sigma,s}} \end{aligned} \tag{62}$$

Remark 3.6. *Note that there are two important aspects of this lemma – the first is that we fix the second argument of the kernel function κ to the value p (or the third to $k - p$) and the second is that the error which we make by this procedure is $\mathcal{O}(\varepsilon)$.*

Proof. We give the details of the proof for the first of the two cases in the Lemma. The very similar second case is left to the reader.

$$\begin{aligned} &\| \int \kappa(\cdot, \cdot - \ell, \ell)\varepsilon^{-1}\widehat{\psi}(\frac{\cdot - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell - \int \kappa(\cdot, p, \ell)\varepsilon^{-1}\widehat{\psi}(\frac{\cdot - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell\|^2_{Y_{\sigma,s}} \\ &= \int \left(\int (\kappa(k, k - \ell, \ell) - \kappa(k, p, \ell))\varepsilon^{-1}\widehat{\psi}(\frac{k - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell \right)^2 e^{2\sigma|k|}(1 + k^2)^s dk \\ &\leq \int \left(C_\kappa \int |(k - \ell) - p|\varepsilon^{-1}\widehat{\psi}(\frac{k - \ell - p}{\varepsilon})\widehat{R}(\ell)d\ell \right)^2 e^{2\sigma|k|}(1 + k^2)^s dk \\ &\leq C_\kappa^2 (\int e^{\sigma m}(1 + m^2)^{s/2}|\frac{m}{\varepsilon}||\widehat{\psi}(\frac{m}{\varepsilon})|dm)^2\|R\|^2_{Y_{\sigma,s}} \leq C_{\psi,\kappa,p}\varepsilon^2\|R\|^2_{Y_{\sigma,s}}, \end{aligned}$$

where to the next to last inequality we applied Young's inequality to bound the L^2-norm of the convolution and the last relied on the fact that $\widehat{\psi}$ has compact support. $\square$

Remark 3.7. *The conclusions of Lemma 3.5 also hold if the integrals run only over a subset of $I\!R$.*

We use Lemma 3.5 to replace the equation (60) with an alternative equation for the components $N^{1,r}_{j,\pm1,n}$ which will result in a form for the normal-form transformation that is easier to bound, at the expense of introducing additional "error" terms all of which are $\mathcal{O}(\varepsilon^2)$. More specifically we apply Lemma 3.5 and make the following changes in (60)

(A.1) We replace $N^{1,r}_{j,\pm1,n}(i\omega\psi_{\pm1}, R^r_n)$ by $N^{1,r}_{j,\pm1,n}(i\omega(\pm k_0)\psi_{\pm1}, R^r_n)$.

(A.2) We replace $\hat{N}^{1,r}_{j,\pm1,n}(\psi_{\pm1}, i\omega_n R^r_n)(k)$ by $\hat{N}^{1,r}_{j,\pm1,n}(\psi_{\pm1}, i\omega_n(k \mp k_0)R^r_n)(k)$.

(A.3) We replace $\hat{b}_{j,\pm1,n}(\psi_{\pm1}, \vartheta_0 R^r_n)(k)$ by $\hat{b}_{j,\pm1,n}(\psi_{\pm1}, \hat{\vartheta}_0(k \mp k_0)R^r_n)(k)$.

Inserting these changes into (60) we find that the kernels of $\hat{N}^{1,r}_{j,\pm1,n}$ should be of the form:

$$\hat{n}^{1,r}_{j,\pm1,n}(k,k-\ell,\ell) \;\; = \;\; \frac{i\widehat{P}^1(k)\hat{b}_{j,\pm1,n}(k,k-\ell,\ell)}{(-\omega_j(k)-\omega(\pm k_0)+\omega_n(k\mp k_0))}\,\frac{\hat{\vartheta}_0(k\mp k_0)}{\hat{\vartheta}(k)}\,. \tag{63}$$

Remark 3.8. *The analysis of the kernel of $N^{0,1}$ would be simplified by the changes (A1)–(A3), too. However, we don't make those changes in $N^{0,1}$ because they would complicate the analysis of the subsequent second normal-form transformation which is required due to the fact that $N^{0,1} = \mathcal{O}(\varepsilon^{-1})$ for certain wave numbers.*

Due to the fact that the support of $\hat{\psi}_{\pm1}$ is non-zero only near $k = \pm k_0$, and the projection operator $\widehat{P}^1$, the expression (63) only has to be analyzed for $|k-\ell\mp k_0| < \delta$ and $|k| \geq \delta$. We now consider the possible resonances in the denominator of (63), taking these restrictions into account.

- $k = 0$: Since $\widehat{P}^1(k) = 0$ for $|k| \leq \delta$, this resonance does not play a role in the analysis of either $N^{1,0}$ or $N^{1,1}$.
- $k = \pm k_0$: The kernels $\hat{n}^{1,r}_{j,\pm1,n}$ have a resonance at $k = \pm k_0$ whenever $j = 2$ and a resonance at $k = \mp k_0$ whenever $j = 3$ and $n = 1$. However, since the derivative of ω_j for $j = 2,3$ at $\pm k_0$ is $\mathcal{O}(1)$, we have a bound on the denominator of the form

  $$|-\omega_j(k)-\omega(\pm k_0)+\omega_n(k\mp k_0)| \geq C|k\mp k_0| \tag{64}$$

  This singularity is offset, however, by the fact that the term $|\hat{\vartheta}_0(k\mp k_0)| \leq C|k\mp k_0|$ and hence the kernels $\hat{n}^{1,r}_{j,\pm1,n}$ can be extended continuously at $k = \pm k_0$ with an $\mathcal{O}(1)$ bound on its size.
- $k = \pm 2k_0$: The kernels $\hat{n}^{1,r}_{j,\pm1,n}$ have a resonance at $k = \pm 2k_0$ whenever $j = 1$ and $n = 3$. However, in this case we have $\hat{b}_{j,\pm1,n}(k,k-\ell,\ell) = \hat{b}_3(k,k-\ell,\ell)$. This will imply that the numerator of $\hat{n}^{1,r}_{j,\pm1,n}$ also vanishes at $k = \pm 2k_0$ so that the quotient is still well-defined. We discuss this surprising **cancellation** phenomenon in detail in Subsection 3.6.

There are no other resonances for the normal-form transforms and hence the kernel can be bounded for all values of k and ℓ by an $\mathcal{O}(1)$ bound.

Having discussed the zeroes of the denominator we are are now interested in the asymptotics for $|k| \to \infty$, in order to see a gain or loss of regularity by the normal-form transform. We first bound the expression for $N^{1,1}_{3,\pm1,3}$. We see from (129) that the numerator of the kernel

$\hat{b}_{3,\pm1,3} = \hat{b}_{13}$ has the form

$$\begin{aligned}
\hat{b}_{13}(k, k-\ell, \ell) \quad = \quad & -\frac{1}{2\hat{s}(k)}\left\{\frac{\hat{s}(k-\ell)\hat{\mathcal{K}}_0(\ell) - \hat{\mathcal{K}}_0(k)\hat{s}(k-\ell)}{\hat{\mathcal{K}}_0(k-\ell)}\right\}(i\ell) \\
& -\hat{s}(k-\ell)\left\{1 + \hat{K}_0(k)\hat{\mathcal{K}}_0(\ell)\right\}(i\ell) \\
& +\frac{1}{2}\ell(k-\ell)\hat{s}(\ell)\hat{s}(k-\ell)\left\{\frac{1}{\hat{\mathcal{K}}_0(k-\ell)} + \hat{\mathcal{K}}_0(\ell)\right\} \\
& -\frac{1}{2\hat{s}(k-\ell)}\left\{\frac{\hat{s}(\ell)\hat{\mathcal{K}}_0(k-\ell) - \hat{\mathcal{K}}_0(k)\hat{s}(\ell)}{\hat{\mathcal{K}}_0(\ell)}\right\} i(k-\ell) \\
& -\hat{s}(\ell)\left\{1 + \hat{\mathcal{K}}_0(k)\hat{\mathcal{K}}_0(k-\ell)\right\} i(k-\ell) \\
& +\frac{1}{2}(k-\ell)\ell\hat{s}(k-\ell)\hat{s}(\ell)\left\{\frac{1}{\hat{\mathcal{K}}_0(\ell)} + \hat{\mathcal{K}}_0(k-\ell)\right\}
\end{aligned} \tag{65}$$

Our goal, throughout the construction of the normal-form transformation is to preserve, as much as possible, the smoothness of the error functions. In practice, this means that when we bound $(1+k^2)^s|\hat{N}^{1,1}(k, k-\ell, \ell)|$ by $C(1+|k-\ell|^2)^{s_1}(1+|\ell|^2)^{s_2}$ in order to apply Lemma 3.1, we will want to keep s_2 small – we can "soak up" as many powers of $|k-\ell|$ as we need since this corresponds to differentiating the approximating function $\psi_{\pm1}$ which is entire. This "trick" of moving derivatives from the factors of w to ψ is related to the smoothing properties of the nonlinear terms in the water wave problem. Thus, for instance, by the estimates in the proof of Corollary 3.13 on page 1499 of [SW00] we see that for all $s > 1$ we have

$$(1+k^2)^{\frac{s}{2}+\frac{1}{4}}\left|\frac{\hat{\mathcal{K}}_0(\ell) - \hat{\mathcal{K}}_0(k)}{\hat{\mathcal{K}}_0(k-\ell)}\right| \leq C(1+|k-\ell|^2)^{\frac{s}{2}+\frac{1}{4}}(1+\ell^2)^{\frac{s-1}{2}} \ . \tag{66}$$

Thus, the expression in the first line on the right hand side of inequality (65) can be bounded by

$$\begin{aligned}
(1+k^2)^{s/2} & \left|\frac{1}{\hat{s}(k)}\left\{\frac{\hat{s}(k-\ell)\hat{\mathcal{K}}_0(\ell) - \hat{\mathcal{K}}_0(k)\hat{s}(k-\ell)}{\hat{\mathcal{K}}_0(k-\ell)}\right\}(i\ell)\right| \\
& \leq (1+k^2)^{\frac{s}{2}+\frac{1}{4}}\left|\frac{\hat{\mathcal{K}}_0(\ell) - \hat{\mathcal{K}}_0(k)}{\hat{\mathcal{K}}_0(k-\ell)}\right||\ell| \leq C(1+|k-\ell|^2)^{\frac{s}{2}+\frac{1}{4}}(1+\ell^2)^{s/2}
\end{aligned} \tag{67}$$

The second line in (65) is bounded by writing

$$1 + \hat{\mathcal{K}}_0(k)\hat{\mathcal{K}}_0(k-\ell) = 1 + (\hat{\mathcal{K}}_0(k))^2 + \hat{\mathcal{K}}_0(k)\left(\hat{\mathcal{K}}_0(\ell) - \hat{\mathcal{K}}_0(k)\right) \ .$$

Noting first that the asymptotics of the hyperbolic tangent imply that $|1+(\hat{\mathcal{K}}_0(k))^2| \leq Ce^{-|k|}$ and using the estimates on page 1499 of [SW00] to bound the term involving $\hat{\mathcal{K}}_0(k) - \hat{\mathcal{K}}_0(\ell)$ we have

$$\begin{aligned}
& (1+k^2)^{s/2}|\hat{s}(k-\ell)||1 + \hat{\mathcal{K}}_0(k)\hat{\mathcal{K}}_0(k-\ell)||\ell| \\
& \qquad \leq C(1+|k-\ell|^2)^{s/2}(1+\ell^2)^{s/2} \ .
\end{aligned} \tag{68}$$

Now, we bound the third line in 65. In this case, unfortunately, there is no cancellation between the factors of $\hat{\mathcal{K}}_0$ and the only smoothing comes from the factors of $\hat{s}(\ell)$ and $\hat{s}(k-\ell)$ so we find

$$\frac{1}{2}(1+k^2)^{s/2}\left|\ell(k-\ell)\hat{s}(\ell)\hat{s}(k-\ell)\left\{\frac{1}{\hat{\mathcal{K}}_0(k-\ell)}+\hat{\mathcal{K}}_0(\ell)\right\}\right| \le C|\ell|^{s+\frac{1}{2}}|k-\ell|^{s+\frac{1}{2}} .$$

Finally, the last three lines can be bounded analogously to the first three lines by changing $k-\ell$ and ℓ.

Combining these estimates on the numerator of the kernel $\hat{n}^{1,1}_{3,\pm1,3}$ with the previous informations of the denominator we find

$$(1+k^2)^{s/2}\left|\hat{n}^{1,1}_{3,\pm1,3}(k,k-\ell,\ell)\right| \le C(1+|\ell|)^{s+\frac{1}{2}}(1+|k-\ell|)^{s+\frac{1}{2}} . \tag{69}$$

From this estimate and Lemma 3.1 we find immediately:

Corollary 3.9. *$N^{1,1}_{3,\pm1,3}$ defines a bounded linear operator from $Y_{\sigma,s+1/2}$ to $Y_{\sigma,s}$.*

We now turn to estimate the terms of all other components of $N^{1,1}$. The terms in the numerators of the kernels can be estimated just as in the case of the terms in the numerator of $\hat{n}^{1,1}_{3,\pm1,3}$ above. More specifically, those terms involving $\mathcal{M}_1$, or $(s\psi+\mathcal{K}_0(s\psi)\mathcal{K}_0)\partial_\alpha$ can be bounded using the smoothing properties exploited on page 1499 of [SW00], so that they result in no loss of smoothness. The terms involving $\mathcal{M}_2$ and $(\partial_\alpha(s\psi))\mathcal{K}_0(\partial_\alpha s\cdot)$ lose half a derivative as above.

Therefore, we find that the normal-form transformation $N^{1,1}$ is well defined, *but causes a loss of smoothness.* That is, we have the following. If $R^1_n \in Y^2_{\sigma,s}$ then there exists $C>0$ such that

$$\|\varepsilon N^{1,1}_{n,\pm1,n}(\psi_{\pm1},R^1_n)\|_{Y^2_{\sigma,s-1/2}} \le C\varepsilon\|R^1_n\|_{Y^2_{\sigma,s}} . \tag{70}$$

Hence, there is a loss of "1/2 a derivative" – i.e., we get a bound of $N^{1,1}$ in the space $Y^2_{\sigma,s-1/2}$ rather than $Y^2_{\sigma,s}$. One the other hand, since we do not have to deal with the large values of $\vartheta^{-1}(k)$ near $k\approx 0$ we obtain a factor of ε on the right–hand side of this estimate.

Taking into account that $|\omega(k)| \sim \sqrt{|k|}$ for $|k|\to\infty$ we further obtain for $j\neq n$ the estimate

$$\|\varepsilon N^{1,1}_{j,\pm1,n}(\psi_{\pm1},R^1_n)\|_{Y^2_{\sigma,s}} \le C\varepsilon\|R^1_n\|_{Y^2_{\sigma,s}} . \tag{71}$$

Moreover, due to the compact support of $\hat{R}^0_n$ the loss of regularity is not present in the estimate for $N^{1,0}$. We find

$$\|\varepsilon N^{1,0}_{j,\pm1,n}(\psi_{\pm1},R^0_n)\|_{Y^2_{\sigma,s}} \le C\varepsilon\|R^0_n\|_{Y^2_{\sigma,s}} . \tag{72}$$

Now, we define the function space

$$Y^R_{\sigma,s} = Y_{\sigma,s}\times Y_{\sigma,s-1/2}\times Y_{\sigma,s-1/2} .$$

Then we can sum up the results of this first normal-form transformation as follows:

Proposition 3.10. *Let $\tilde{R}^r$ for $r=0,1$ be defined by (52). Then this transformation maps $(R^0,R^1)\in Y^R_{\sigma,s}\times Y^R_{\sigma,s}$ into $(\tilde{R}^0,\tilde{R}^1)\in Y^R_{\sigma,s}\times Y^R_{\sigma,s-1/2}$ for all $s>1$ and $\sigma\ge 0$ and is invertible on its range. Furthermore, if we write the inverse transformations as*

$$R^0 = \tilde{R}^0 + \mathcal{N}_0^{-1}(\tilde{R}^0,\tilde{R}^1) , \qquad R^1 = \tilde{R}^1 + \mathcal{N}_1^{-1}(\tilde{R}^0,\tilde{R}^1) ,$$

then there exist constants C_0, C_1 such that the inverse transformations satisfy the estimates

$$\begin{aligned}\|\mathcal{N}_0^{-1}(\tilde{R}^0,\tilde{R}^1)\|_{Y^R_{\sigma,s}} &\leq C_0(\|\tilde{R}^0\|_{Y^R_{\sigma,s}}+\|\tilde{R}^1\|_{Y^R_{\sigma,s}})\ ,\\ \|\mathcal{N}_1^{-1}(\tilde{R}^0,\tilde{R}^1)\|_{Y^R_{\sigma,s}} &\leq C_1\varepsilon(\|\tilde{R}^0\|_{Y^R_{\sigma,s}}+\|\tilde{R}^1\|_{Y^R_{\sigma,s}})\,.\end{aligned}$$

Finally, if (R^0,R^1) satisfy the equations (51) then $(\tilde{R}^0,\tilde{R}^1)$ satisfy

$$\begin{aligned}\partial_t\tilde{R}^0_j &= i\omega_j\tilde{R}^0_j+\varepsilon\sum_{\substack{l_1,l_2=-1,1\\ m=1,2,3\\ n=1,2,3}} N^{0,1}_{j,l_1,m}(\psi_{l_1},\varepsilon\vartheta^{-1}P^1B_{m,l_2,n}(\psi_{l_2},\vartheta\tilde{R}^1_n))+\varepsilon^2\mathcal{E}^0_j\ , \qquad (73)\\ \partial_t\tilde{R}^1_j &= \omega_j\tilde{R}^1_j+\varepsilon^2\mathcal{E}^1_j\ ,\end{aligned}$$

for $j=1,2,3$, where $\varepsilon^2\mathcal{E}^r=\varepsilon^2(\mathcal{E}^r_1,\mathcal{E}^r_2,\mathcal{E}^r_3)$ for $r=0,1$ denotes a collection of terms whose $Y^R_{\sigma,s-1}$-norms are bounded by $C\varepsilon^2$ for $(\tilde{R}^0,\tilde{R}^1)$ in some fixed ball in $Y^R_{\sigma,s}\times Y^R_{\sigma,s}$.

Proof. The proof of invertibility of the transformation is deferred until the next section. Assuming the invertibility for the moment the structure of the equations (73) follows immediately using $R^1_j=\tilde{R}^1_j+\mathcal{O}(\varepsilon)$ for $j=1,2,3$. □

3.5 Properties of the nonlinear terms for $k\to 0$

This subsection, along with the following two, give details about special features of the terms that appear in the normal-form transformation which we have used in the construction in the preceding section.

We start with some estimates for the nonlinear terms for $k\to 0$. Let $\mathcal{B}_j$ be the kernel of $B_j(f,g)+B_j(g,f)$, i.e.,

$$\mathcal{F}(B_j(f,g)+B_j(g,f))(k) = \int\hat{\mathcal{B}}_j(k,k-\ell,\ell)\hat{f}(k-\ell)\hat{g}(\ell)d\ell.$$

Then we have

Lemma 3.11. $|\hat{\mathcal{B}}_j(k,k-\ell,\ell)|\leq C|k|$ *for $j=4,9,14$.*

Proof. Due to the smoothness of kernel it is sufficient to show $\hat{\mathcal{B}}_j(0,-\ell,\ell)=0$. Consider first

$$\begin{aligned}B_4(f,g)+B_4(g,f) &= -\mathcal{M}_1(sf,\partial_\alpha g)-\mathcal{M}_1(sg,\partial_\alpha f)\\ &\quad+(sf)(\partial_\alpha g)+(sg)(\partial_\alpha f)\\ &\quad+\mathcal{K}_0((sf)\mathcal{K}_0(\partial_\alpha g))+\mathcal{K}_0((sg)\mathcal{K}_0(\partial_\alpha f)).\end{aligned}$$

1. We start by considering the term in the second line above, i.e.,

$$\begin{aligned}&\int\hat{s}(k-\ell)\hat{f}(k-\ell)i\ell\hat{g}(\ell)+\hat{s}(k-\ell)\hat{g}(k-\ell)i\ell\hat{f}(\ell)d\ell\\ =&\int\hat{s}(k-\ell)\hat{f}(k-\ell)i\ell\hat{g}(\ell)+\hat{s}(\ell)\hat{g}(\ell)i(k-\ell)\hat{f}(k-\ell)d\ell.\end{aligned}$$

The kernel

$$\hat{s}(k-\ell)i\ell+\hat{s}(\ell)i(k-\ell)$$

vanishes for $k=0$ using that $\hat{s}$ is an even function.

2. If we now check the third line we have

$$\mathcal{K}_0(k)\int \hat{s}(k-\ell)\hat{f}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{g}(\ell)+\hat{s}(k-\ell)\hat{g}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{f}(\ell)d\ell.$$

   Since $|\mathcal{K}_0(k)|\leq|k|$ we are also done with this term.

3. Finally we come to the first line, namely

$$\begin{aligned}\int(&-\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{f}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{g}(\ell)\\ &+\mathcal{K}_0(k)\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{f}(k-\ell)i\ell\hat{g}(\ell)\\ &-\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{g}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{f}(\ell)\\ &+\mathcal{K}_0(k)\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{g}(k-\ell)i\ell\hat{f}(\ell))d\ell.\end{aligned}$$

   Using that $\hat{s}$ is even and $\mathcal{K}_0$ is odd shows that the kernel

$$\frac{\mathcal{K}_0(k)-\mathcal{K}_0(\ell)}{\mathcal{K}_0(k-\ell)}\hat{s}(k-\ell)i\ell+\frac{\mathcal{K}_0(k)-\mathcal{K}_0(k-\ell)}{\mathcal{K}_0(\ell)}\hat{s}(\ell)i(k-\ell)$$

   vanishes for $k=0$.

The next term to consider is $B_9(f,g)+B_9(g,f)$. Many of the terms in this expression appeared above but the new ones we consider are

$$\begin{aligned}&-\mathcal{M}_2(sf)\partial_\alpha(sg)-\mathcal{M}_2(sg)\partial_\alpha(sf)\\ &+\partial_\alpha(sf)\mathcal{K}_0\partial_\alpha(sg)+\partial_\alpha(sg)\mathcal{K}_0\partial_\alpha(sf)\\ =&\int-\mathcal{M}_2(k-\ell)\hat{s}(k-\ell)\hat{f}(k-\ell)i\ell\hat{s}(\ell)\hat{g}(\ell)\\ &-\mathcal{M}_2(k-\ell)\hat{s}(k-\ell)\hat{g}(k-\ell)i\ell\hat{s}(\ell)\hat{f}(\ell)\\ &+i(k-\ell)\hat{s}(k-\ell)\hat{f}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{s}(\ell)\hat{g}(\ell)\\ &+i(k-\ell)\hat{s}(k-\ell)\hat{g}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{s}(\ell)\hat{f}(\ell)d\ell\ .\end{aligned}$$

Using that $\hat{s}$ and $\mathcal{M}_2$ are even and $\mathcal{K}_0$ is odd shows that the kernels

$$-\mathcal{M}_2(k-\ell)\hat{s}(k-\ell)i\ell\hat{s}(\ell)-\mathcal{M}_2(\ell)\hat{s}(\ell)i(k-\ell)\hat{s}(k-\ell)$$

and

$$i(k-\ell)\hat{s}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{s}(\ell)+i(\ell)\hat{s}(\ell)\mathcal{K}_0(k-\ell)i(k-\ell)\hat{s}(k-\ell)$$

vanish for $k=0$.

The term $B_{14}(f,g)+B_{14}(g,f)$ works analogously because all types of terms in this expression have already appeared above. □

3.6 Cancellation

In the following we discuss the terms indicated before with cancellation. We have to show the cancellation of $B_3(\psi_{\pm1},\vartheta R_3)$ at the wave number $k=\pm2k_0$ in the equation for R_1. We need the boundedness of

$$\frac{\hat{b}_3(k,k-\ell,\ell)}{-\omega(k-\ell)+\omega(\ell)}$$

for $|k-\ell \mp k_0| \leq \delta$. The denominator vanishes for the terms under consideration for $k=2\ell$. Hence we have to prove that also

$$\hat{b}_3(2\ell,\ell,\ell) = 0 \; . \tag{74}$$

Due to the scalings the relevant quadratic terms on the right-hand side in the equation for c_1 are

$$\begin{aligned}
&-\mathcal{M}_1(sc_2,\partial_\alpha c_3) + \mathcal{M}_1(sc_3,\partial_\alpha c_2)\\
&+sc_2\partial_\alpha c_3 - sc_3\partial_\alpha c_2\\
&+\mathcal{K}_0(sc_2\mathcal{K}_0))\partial_\alpha c_3 - \mathcal{K}_0(sc_3\mathcal{K}_0))\partial_\alpha c_2 .
\end{aligned}$$

1. We consider the second line first. We plug in the ansatz and find

$$\int \hat{\psi}_{\pm 1}(k-\ell)\hat{s}(k-\ell)i\ell\hat{\vartheta}(\ell)\hat{R}_3(\ell) - \hat{\vartheta}(k-\ell)\hat{R}_3(k-\ell)\hat{s}(k-\ell)i\ell\hat{\psi}_{\pm 1}(\ell)d\ell .$$

   We make a coordinate transform $\ell \to k-\ell$ in the second part of the integral. Then the kernel

$$\hat{s}(k-\ell)i\ell - \hat{s}(\ell)i(k-\ell)$$

   vanishes for $k=2\ell$ as stated above.

2. Next we consider the third line. We plug in the ansatz and find

$$\begin{aligned}
\int &\mathcal{K}_0(k)\hat{\psi}_{\pm 1}(k-\ell)\hat{s}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{\vartheta}(\ell)\hat{R}_3(\ell)\\
&-\mathcal{K}_0(k)\hat{\vartheta}(k-\ell)\hat{R}_3(k-\ell)\hat{s}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{\psi}_{\pm 1}(\ell)d\ell .
\end{aligned}$$

   We make a coordinate transform $\ell \to k-\ell$ in the second part of the integral. Then the kernel

$$\mathcal{K}_0(k)\hat{s}(k-\ell)\mathcal{K}_0(\ell)i\ell - \mathcal{K}_0(k)\hat{s}(\ell)\mathcal{K}_0(k-\ell)i(k-\ell)$$

   vanishes for $k=2\ell$ as stated above.

3. Finally consider the first line. Recall that

$$\begin{aligned}
\mathcal{M}_1(Z_1,U_1) &= [\mathcal{K}_0^{-1}Z_1,\mathcal{K}_0]U_1\\
&= (\mathcal{K}_0^{-1}Z_1)(\mathcal{K}_0U_1) - \mathcal{K}_0((\mathcal{K}_0^{-1}Z_1)U_1).
\end{aligned}$$

   Thus we get

$$\begin{aligned}
&-\mathcal{M}_1(s\psi_{\pm 1},\partial_\alpha(\vartheta R_3)) + \mathcal{M}_1(s\vartheta R_3,\partial_\alpha\psi_{\pm 1})\\
=\;&-(\mathcal{K}_0^{-1}(s\psi_{\pm 1}))(\mathcal{K}_0\partial_\alpha(\vartheta R_3)) + \mathcal{K}_0((\mathcal{K}_0^{-1}(s\psi_{\pm 1}))\partial_\alpha(\vartheta R_3))\\
&+(\mathcal{K}_0^{-1}(s\vartheta R_3))(\mathcal{K}_0\partial_\alpha\psi_{\pm 1}) - \mathcal{K}_0((\mathcal{K}_0^{-1}(s\vartheta R_3))\partial_\alpha\psi_{\pm 1})\\
=\;&\int -\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{\psi}_{\pm 1}(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{\vartheta}(\ell)\hat{R}_3(\ell)\\
&+\mathcal{K}_0(k)\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{\psi}_{\pm 1}(k-\ell)i\ell\hat{\vartheta}(\ell)\hat{R}_3(\ell)\\
&+\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{\vartheta}(k-\ell)\hat{R}_3(k-\ell)\mathcal{K}_0(\ell)i\ell\hat{\psi}_{\pm 1}(\ell)\\
&-\mathcal{K}_0(k)\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\hat{\vartheta}(k-\ell)\hat{R}_3(k-\ell)i\ell\hat{\psi}_{\pm 1}(\ell))d\ell \, .
\end{aligned}$$

We make a coordinate transform $\ell \to k-\ell$ in the second part of the integral. Then the kernel

$$\begin{aligned}
&-\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)\mathcal{K}_0(\ell)i\ell \\
&+\mathcal{K}_0(\ell)\mathcal{K}_0^{-1}(k-\ell)\hat{s}(k-\ell)i\ell \\
&+\mathcal{K}_0^{-1}(\ell)\hat{s}(\ell)\mathcal{K}_0(k-\ell)i(k-\ell) \\
&-\mathcal{K}_0(k)\mathcal{K}_0^{-1}(\ell)\hat{s}(\ell)i(k-\ell)
\end{aligned}$$

vanishes for $k=2\ell$ as stated above. The first and the third line, and the second and the fourth line cancel.

Hence, we have shown (74).

3.7 Long wave form

We now verify that as we claimed just after (51), the terms in the third line of the equation for R_j^0 are $\mathcal{O}(\varepsilon^2)$, and hence can be ignored. First, we address the terms $\varepsilon^2\vartheta^{-1}P^0B_{j,0,n}(\Psi_0,\vartheta R_n^r)$ for $j,n=1,2,3$ and $r=0,1$. We split $B_{j,0,n}(\Psi_0,\vartheta R_n^r)$ in the components

$$B_{j,0,n}(\Psi_0,\vartheta R_n^r) = \sum_{i=1,2,3} B_{j,0,n}^i(\psi_{0i},\vartheta R_n^r)\ .$$

By applying the same methods as we used to bound $\hat{b}_{3,\pm1,3}$ as well as the inequality $|\ell| \leq |k|+|k-\ell|$ we get

$$\sup_{|\ell|=\mathcal{O}(1)} \left|(\hat{P}^0\hat{b}_{j,0,n}^i)(k,k-\ell,\ell)\right| \leq C(|k|+|k-\ell|) \tag{75}$$

for $i=1,2,3$, where $\hat{b}_{j,0,n}^i(k,k-\ell,\ell)$ is the kernel of $B_{j,0,n}^i$. This estimate is a consequence of the fact that each summand in $B_{j,0,n}$ contains at least one α-derivative. Using this estimate and the bound $|k|/(\varepsilon+|k|)\leq 1$ we obtain

$$\begin{aligned}
&\left|\int \hat{P}^0(k)\hat{\vartheta}^{-1}(k)\hat{b}_{j,0,n}^i(k,k-\ell,\ell)\hat{\vartheta}(\ell)\hat{\psi}_{0i}(k-\ell)\hat{R}_n^r(\ell)d\ell\right| \\
\leq\ & C\left(\int |\hat{P}^0(k)||k\hat{\vartheta}^{-1}(k)||\hat{\psi}_{0i}(k-\ell)||\hat{R}_n^r(\ell)|d\ell + \int |\hat{P}^0(k)||\tfrac{k-\ell}{\varepsilon}||\hat{\psi}_{0i}(k-\ell)||\hat{R}_n^r(\ell)|d\ell\right) \\
\leq\ & C\left(\int |\hat{P}^0(k)||\hat{\psi}_{0i}(k-\ell)||\hat{R}_n^r(\ell)|d\ell + \int |\hat{P}^0(k)||\widehat{\partial_X\psi_{0i}}(k-\ell)||\hat{R}_n^r(\ell)|d\ell\right)
\end{aligned}$$

for $i=1,2,3$. This implies that for any σ', s' we have

$$\|\varepsilon^2\vartheta^{-1}P^0B_{j,0,n}(\Psi_0,\vartheta R_n^r)\|_{Y_{\sigma',s'}} \leq \varepsilon^2 C_{\sigma',s'}(\Psi_0)\|R_n^r\|_{Y_{\sigma,s}}\ , \tag{76}$$

where the constant $C_{\sigma',s'}(\Psi_0)$ is independent of ε due to the fact that Ψ_0 is long wave-length, i.e., the Fourier transform of Ψ_0 is strongly concentrated near $k=0$. Hence, the terms $\varepsilon^2\vartheta^{-1}P^0B_{j,0,n}(\Psi_0,\vartheta R_n^r)$ are of order $\mathcal{O}(\varepsilon^2)$ and need not to be eliminated.

Finally, we turn to the consideration of the trilinear terms $\varepsilon^2\vartheta^{-1}P^0\mathcal{T}_{j,1,-1,n}(\psi_1,\psi_{-1},\vartheta R_n^r)$ for $j,n=1,2,3$ and $r=0,1$. Using [Cr85, Lemma 3.7] and [SW00, Corollary 3.16] as well as the inequality $|p|\leq|k|+|k-p|$ we get

$$\sup_{|p|=\mathcal{O}(1)} \left|(\hat{P}^0\hat{\mathcal{T}}_{j,1,-1,n})(k,k-\ell,\ell-p,p)\right| \leq C(|k|+|k-p|)\ , \tag{77}$$

where $\hat{\mathcal{T}}_{j,1,-1,n}(k, k-\ell, \ell-p, p)$ is the kernel of $\mathcal{T}_{j,1,-1,n}$. This estimate is a consequence of the fact that each summand in $\mathcal{T}_{j,1,-1,n}$ contains at least one α-derivative. Using this estimate and the bound $|k|/(\varepsilon + |k|) \leq 1$ we obtain

$$\begin{aligned}
& \left| \int \hat{P}^0(k)\vartheta^{-1}(k)\hat{\mathcal{T}}_{j,n}(k, k-\ell, \ell-p, p)\hat{\vartheta}(p)\hat{\psi}_1(k-\ell)\hat{\psi}_{-1}(\ell-p)\hat{R}_n^r(p)d\ell dp \right| \\
\leq \;& C\int |\hat{P}^0(k)|\,|\hat{\psi}_1(k-\ell)|\,|\hat{\psi}_{-1}(\ell-p)|\,|\hat{R}_n^r(p)|\,d\ell dp \\
& + C\int |\hat{P}^0(k)|\,|\tfrac{k-p}{\varepsilon}|\,|\hat{\psi}_1(k-\ell)|\,|\hat{\psi}_{-1}(\ell-p)|\,|\hat{R}_n^r(p)|\,d\ell dp \; .
\end{aligned}$$

This implies that for any σ', s' we have

$$\|\varepsilon^2\vartheta^{-1}P^0\mathcal{T}_{j,1,-1,n}(\psi_1, \psi_{-1}, \vartheta R_n^r)\|_{Y_{\sigma',s'}} \leq \varepsilon^2 C_{\sigma',s'}(\psi_1, \psi_{-1})\|R_n^r\|_{Y_{\sigma,s}} \; , \tag{78}$$

where the constant $C_{\sigma',s'}(\psi_0)$ is independent of ε due to the fact that the Fourier transform of $\psi_1\psi_{-1}$ is strongly concentrated near 0. Hence, the terms $\varepsilon^2\vartheta^{-1}P^0\mathcal{T}_{j,1,-1,n}(\psi_1, \psi_{-1}, \vartheta R_n^r)$ are also of order $\mathcal{O}(\varepsilon^2)$ and need not to be eliminated.

3.8 The second normal-form transform

We now construct a second normal-form transformation to remove the remaining terms of $\mathcal{O}(\varepsilon)$ from (73). Before doing so we analyze the offending terms in more detail. The terms can be written as

$$\begin{aligned}
& \varepsilon\hat{N}^{0,1}_{j,l_1,m}(\psi_{l_1}, \varepsilon\vartheta^{-1}P^1B_{m,l_2,n}(\psi_{l_2}, \vartheta\tilde{R}_n^1))(k) \\
& \quad = \varepsilon^2\int \hat{n}^{0,1}_{j,l_1,m}(k, k-\ell, \ell)\hat{\psi}_{l_1}(k-\ell) \\
& \qquad \times\vartheta^{-1}(\ell)\widehat{P}^1(\ell)\left(\int \hat{b}_{m,l_2,n}(\ell, \ell-p, p)\hat{\psi}_{l_2}(\ell-p)\hat{\vartheta}(p)\hat{\tilde{R}}_n^1(p)dp\right)d\ell \; ,
\end{aligned} \tag{79}$$

where we recall that

$$\hat{n}^{0,1}_{j,l_1,m}(k, k-\ell, \ell) = \frac{i\widehat{P}^0(k)\hat{b}_{j,l_1,m}(k, k-\ell, \ell)}{(-\omega_j(k) - \omega(k-\ell) + \omega_n(\ell))}\frac{\hat{\vartheta}(\ell)}{\hat{\vartheta}(k)}.$$

We now apply Lemma 3.5 to simplify this expression as we did for $N^{1,r}$ $(r = 0, 1)$. If we do so we obtain the expression

$$\begin{aligned}
& \varepsilon\hat{N}^{0,1}_{j,l_1,m}(\psi_{l_1}, \varepsilon\vartheta^{-1}(\cdot - l_1k_0)P^1B_{m,l_2,n}(\psi_{l_2}, \vartheta(\cdot - l_2k_0)\tilde{R}_n^1))(k) \\
& \quad = \varepsilon^2\int \hat{\hat{n}}^{0,1}_{j,l_1,m}(k)\hat{\psi}_{l_1}(k-\ell)\widehat{P}^1(k - l_1k_0) \\
& \quad \times\left(\int \hat{b}_{m,l_2,n}(k - l_1k_0, k - l_1k_0 - p, p)\hat{\psi}_{l_2}(\ell-p)\hat{\vartheta}(k - (l_1+l_2)k_0)\hat{\tilde{R}}_n^1(p)dp\right)d\ell \\
& \quad +\varepsilon^2\mathcal{E}^{0,1}_j \; ,
\end{aligned} \tag{80}$$

where $\varepsilon^2\mathcal{E}^{0,1} = \varepsilon^2(\mathcal{E}^{0,1}_1, \mathcal{E}^{0,1}_2, \mathcal{E}^{0,1}_3)$ denotes a collection of terms whose $Y^R_{\sigma,s-1}$-norms are bounded by $C\varepsilon^2$ for $(\tilde{R}^0, \tilde{R}^1)$ in some fixed ball in $Y^R_{\sigma,s} \times Y^R_{\sigma,s}$.

Moreover, we use the abbreviation

$$\hat{\hat{n}}^{0,1}_{j,l_1,m}(k) = \frac{i\widehat{P}^0(k)\hat{b}_{j,l_1,m}(k, l_1k_0, k-l_1k_0)}{(-\omega_j(k) - \omega(l_1k_0) + \omega_m(k-l_1k_0))} \frac{1}{\hat{\vartheta}(k)}.$$

With these modifications we can now prove that all terms of the form (79) with $l_1 = -l_2$ are $\mathcal{O}(\varepsilon^2)$ and hence can be included in the $\varepsilon^2\mathcal{E}^r_j$ terms in (73):

Lemma 3.12. *There exists $C > 0$ such that*

$$\begin{array}{rcl} \|\varepsilon N^{0,1}_{j,1,m}(\psi_1, \varepsilon\vartheta^{-1}P^1B_{m,-1,n}(\psi_{-1}, \vartheta\tilde{R}^1_n))\|_{Y_{\sigma,s}} & \leq & C\varepsilon^2\|\tilde{R}^1_n\|_{Y_{\sigma,s}}, \\ \|\varepsilon N^{0,1}_{j,-1,m}(\psi_{-1}, \varepsilon\vartheta^{-1}P^1B_{m,1,n}(\psi_1, \vartheta\tilde{R}^1_n))\|_{Y_{\sigma,s}} & \leq & C\varepsilon^2\|\tilde{R}^1_n\|_{Y_{\sigma,s}}. \end{array}$$

Proof. Since $N^{0,1}_{j,\pm1,m}$ contains the factor $\widehat{P}^0(k)$ means that the integral over k which occurs in the $Y_{\sigma,s}$-norm runs only over the integral $|k| < \delta$. Thus, we can bound the $Y_{\sigma,s}$-norm by bounding the maximum of the kernel. The first term in Lemma 3.12 has the modified kernel

$$\varepsilon^2\hat{\hat{n}}_{j,1,m}(k)\widehat{P}^1(k-k_0)\hat{b}_{m,-1,n}(k-k_0, k-k_0-p, p)\widehat{\vartheta}(k). \tag{81}$$

Since $\vartheta(k)\hat{\hat{n}}_{j,1,m}(k)$ is $\mathcal{O}(1)$ bounded and all other terms in (81) are $\mathcal{O}(1)$ bounded for $|k| < \delta$ we have an $\mathcal{O}(\varepsilon^2)$ bound for the kernel (81). The second term in Lemma 3.12 can be estimated similarly. □

Lemma 3.12 implies that the terms of the form (79) with $l_1 = -l_2$ need not be eliminated by the normal-form transformation. Thus we now turn to the terms of the form (79) with $l_1 = l_2$. If we simplify the kernels of these terms with the aid of Lemma 3.5, we find the kernels have the form:

$$\varepsilon^2\hat{\hat{n}}_{j,\pm1,m}(k)\widehat{P}^1(k \mp k_0)\hat{b}_{m,\pm1,n}(k \mp k_0, k \mp k_0 - p, p)\widehat{\vartheta}(k \mp 2k_0) \tag{82}$$

plus errors that are of size $\mathcal{O}(\varepsilon^2)$. Note that in contrast to the terms considered in Lemma 3.12 this expression does not contain a factor of $\hat{\vartheta}(k)$ to offset the $\hat{\vartheta}(k)$ in the denominator of $\hat{\hat{n}}_{j,\pm1,m}(k)$ and thus they must be eliminated by a second normal-form transformation.

We look for a transformation of the form

$$\begin{array}{rcl} \mathcal{R}^0_j & = & \tilde{R}^0_j + \varepsilon D^{0,1,+}_j(\psi_1, \psi_1, \tilde{R}^1) + \varepsilon D^{0,1,-}_j(\psi_{-1}, \psi_{-1}, \tilde{R}^1) \\ \mathcal{R}^1_j & = & \tilde{R}^1_j\ . \end{array} \tag{83}$$

Differentiating the expression for $\mathcal{R}^0_j$ we find, just as in Subsection 3.2, that the terms of $\mathcal{O}(\varepsilon)$ in (73) will be eliminated if $D^{0,1,+}_j$ satisfies

$$\begin{aligned} \Big\{ & -i\omega_j D^{0,1,+}_j(\psi_1, \psi_1, \tilde{R}^1) - D^{0,1,+}_j(i\omega\psi_1, \psi_1, \tilde{R}^1) - D^{0,1,+}_j(\psi_1, i\omega\psi_1, \tilde{R}^1) \\ & + D^{0,1,+}_j(\psi_1, \psi_1, \Lambda\tilde{R}^1) + N^{0,1}_j(\psi_1, \varepsilon\vartheta^{-1}P^1B(\psi_1, \vartheta\tilde{R}^1)) \Big\} = 0\,, \end{aligned} \tag{84}$$

where $\hat{\Lambda}(k)$ is a diagonal matrix with entries $i\omega_j(k)$, $j = 1, 2, 3$. We find that we have to choose

$$\begin{aligned} & \varepsilon D^{0,1,+}_j(\psi_1, \psi_1, \tilde{R}^1) \\ & \quad = \varepsilon^2 \sum_{m=1,2,3} \int \hat{\hat{n}}_{j,1,m}(k)\hat{\psi}_1(k-\ell)\widehat{P}^1(k-k_0) \\ & \qquad \times \left(\sum_{n=1,2,3} \int \frac{\hat{b}_{m,1,n}(k-k_0, k-k_0-p, p)\hat{\psi}_1(\ell-p)\hat{\vartheta}(k-2k_0)\hat{\tilde{R}}^1_n(p)}{-\omega_j(k) - 2\omega(k_0) + \omega_n(k-2k_0)} dp \right) d\ell\,, \end{aligned} \tag{85}$$

where we used as above in the kernel that $k-\ell \approx \ell - p \approx k_0$ due to the localization of $\hat{\psi}_1$ so we have $p \approx -2k_0$ which is made rigorous with Lemma 3.5. According to Young's inequality we have to estimate the kernel w.r.t. the sup norm. We already know that the numerator in this expression in $\mathcal{O}(\varepsilon)$. In order to estimate the denominator note that in this expression $k \approx 0$ due to the factor of $\widehat{P}^0$ in $\hat{\tilde{n}}_{j,1,m}(k)$. Hence

$$(-\omega_j(k) - 2\omega(k_0) + \omega_n(k-2k_0)) \approx -2\omega(k_0) - \omega_n(2k_0) \neq 0.$$

Regardless of the value of j and n this expression is bounded strictly away from zero. Hence the mapping $\varepsilon D^{0,1,+}$ is $\mathcal{O}(\varepsilon)$-bounded. We can construct and estimate an analogous expression for $D^{0,1,-}$ in a very similar fashion. Therefore, the normal-form transform is well defined and invertible. We find

Lemma 3.13. *If*

$$\mathcal{R}^0 = \tilde{R}^0 + \varepsilon D^{0,1,+}(\psi_1, \psi_1, \tilde{R}^1) + \varepsilon D^{0,1,-}(\psi_{-1}, \psi_{-1}, \tilde{R}^1)$$

with $\varepsilon D^{0,1,\pm}$ defined as in (85)*, then for any $\sigma \geq 0$ and $s > 1$ there exists $C > 0$ such that*

$$\|\varepsilon D^{0,1,\pm}(\psi_{\pm 1}, \psi_{\pm 1}, \tilde{R}^1)\|_{Y^R_{\sigma,s}} \leq C\varepsilon \|\tilde{R}^1\|_{Y^R_{\sigma,s}} .$$

Remark 3.14. *Note that there is no loss of smoothness in this transformation due to the factor of $\widehat{P}^0$ in* (85) *via $\hat{\tilde{n}}_{j,\pm 1,m}(k)$.*

Now, just as in Proposition 3.10 we have:

Proposition 3.15. *Fix $\sigma \geq 0$ and $s \geq 1$. Suppose $(\tilde{R}^0, \tilde{R}^1)$ satisfy the equations* (73)*. Define $(\mathcal{R}^0, \mathcal{R}^1)$ via the transformations* (83)*. Then for any $\rho > 0$, there exists $\varepsilon_\rho > 0$ such that for all $|\varepsilon| < \varepsilon_\rho$ the transformation* (83) *is invertible on the ball of radius ρ in $Y^R_{\sigma,s} \times Y^R_{\sigma,s}$. Furthermore, $(\mathcal{R}^0, \mathcal{R}^1)$ satisfy the equations*

$$\begin{aligned} \partial_t \mathcal{R}^0 &= \Lambda \mathcal{R}^0 + \varepsilon^2 \mathcal{E}^0 , \\ \partial_t \mathcal{R}^1 &= \Lambda \mathcal{R}^1 + \varepsilon^2 \mathcal{E}^1 , \end{aligned} \tag{86}$$

where $\hat{\Lambda}(k)$ is a diagonal matrix with entries $i\omega_j(k)$, $j = 1, 2, 3$ and $\varepsilon^2 \mathcal{E}^r$, $r = 0, 1$, denotes a collection of terms whose $Y^R_{\sigma,s-1}$ norms are bounded by $C\varepsilon^2$.

Proof. The invertibility of the transformation in this case results from a simple application of the Neumann series since there is no loss of smoothness. The equation for $\mathcal{R}^0$ and $\mathcal{R}^1$ follow in the same way the equations for $\tilde{R}^0$ and $\tilde{R}^1$ were derived in the proof of Proposition 3.10. ∎

Finally, we consider the composition of the two normal-form transformations, namely

$$\begin{aligned} \mathcal{R}^0 &= \tilde{R}^0 + \varepsilon D^{0,1,+}(\psi_1, \psi_1, \tilde{R}^1) + \varepsilon D^{0,1,-}(\psi_{-1}, \psi_{-1}, \tilde{R}^1) \\ &= R^0 + \varepsilon N^{0,1}(\Psi, R^1) + \varepsilon D^{0,1,+}(\psi_1, \psi_1, R^1 + \varepsilon N^{1,0}(\Psi, R^0) + \varepsilon N^{1,1}(\Psi, R^1)) \\ &\qquad + \varepsilon D^{0,1,-}(\psi_{-1}, \psi_{-1}, R^1 + \varepsilon N^{1,0}(\Psi, R^0) + \varepsilon N^{1,1}(\Psi, R^1)) \\ &\equiv R^0 + \varepsilon F^0(R) , \end{aligned} \tag{87}$$

with a similar expression for $\mathcal{R}^1 \equiv R^1 + \varepsilon F^1(R)$. From Proposition 3.10 and Proposition 3.15 we see that

1. F^0 and F^1 are linear functions of R, and

2. The (composite) normal-form transformation loses at most half a derivative, i.e., there exists a constant C_F such that

$$\|\varepsilon F^1(R)\|_{Y^R_{\sigma,s-1/2}} \leq C_F\varepsilon\|R\|_{Y^R_{\sigma,s}} .$$

   There is no loss of regularity in F^0 due to its compact support in Fourier space.

If we now insert the information we have derived on the equations satisfied by the transformed variables we find the following proposition:

Proposition 3.16. *There exists a (linear) change of variables,*

$$\mathcal{R} = R + \varepsilon F(R)$$

defined for $R \in Y^R_{\sigma,s}$ and invertible on its range such that in terms of the transformed variables the equation for the evolution of the error in our approximation takes the form

$$\partial_t \mathcal{R} = \Lambda\mathcal{R} + \varepsilon^2\ell(\mathcal{R}) + \varepsilon^3 G(\mathcal{R}) + \varepsilon^{-3}\vartheta^{-1}\mathrm{Res}(\varepsilon\Psi) . \tag{88}$$

Furthermore the linear term $\varepsilon^2\ell(\mathcal{R})$ and the bilinear term $\varepsilon^3 G(\mathcal{R})$ satisfy the estimates

$$\|\varepsilon^2\ell(\mathcal{R})\|_{Y^R_{\sigma,s-1}} \leq C_L\varepsilon^2\|\mathcal{R}\|_{Y^R_{\sigma,s}} ,$$

and

$$\|\varepsilon^3 G(\mathcal{R})\|_{Y^R_{\sigma,s-1}} \leq C_G\varepsilon^3\|\mathcal{R}\|_{Y^R_{\sigma,s}}\|\mathcal{R}\|_{Y^R_{\sigma,s-1}} .$$

Proof. The proof follows from the estimates in Proposition 3.10 and Proposition 3.15. The last estimate also relies on the estimates in [SW00, Lemma 3.14, Lemma 3.15 and Corollary 3.16] which exclude the occurrence of terms quadratic in $\|\mathcal{R}\|_{Y^R_{\sigma,s}}$. □

4 Inverting the normal-form transform

To complete the derivation of the evolution equation for $(\mathcal{R}^0, \mathcal{R}^1)$ in Proposition 3.15 we now prove the invertibility of the first normal-form transform asserted in Proposition 3.10. There is a serious problem due to fact that $N^{1,1}$ loses half a derivative, i.e., is a mapping from $Y^R_{\sigma,s}$ into $Y^R_{\sigma,s-\frac{1}{2}}$. Therefore, inverting the normal-form transform with the help of Neumann's series is not possible.

The basic idea behind the inversion is the use of energy estimates to invert the transformation. In the following we explain this strategy by reviewing the handling of a model problem from [SW11, Section 5].

We consider a linear transformation $v \mapsto u$ which can be written in Fourier variables as:

$$\hat{u}(k) = \hat{v}(k) + \varepsilon \int \hat{b}(k)\hat{a}(k-m)\hat{v}(m)dm.$$

We assume that:

- $\widehat{b}(k)$ is pure imaginary.
- $\widehat{b}(k)$ is Lipschitz as a function of k with a Lipschitz constant which is independent of $0 < \varepsilon \ll 1$.

- $\widehat{b}(k) \sim ik$ for $|k| \to \infty$

Furthermore, a is assumed to be smooth and real-valued.

Then we find

$$\begin{aligned}
\int \overline{\hat{v}(k)}\hat{u}(k) + \hat{v}(k)\overline{\hat{u}(k)} &= 2\int \overline{\hat{v}(k)}\hat{v}(k) + \varepsilon \int \overline{\hat{v}(k)}\hat{b}(k)\hat{a}(k-m)\hat{v}(m)dm\,dk \\
&\quad +\varepsilon \int \hat{v}(k)\overline{\hat{b}(k)}\;\overline{\hat{a}(k-m)}\;\overline{\hat{v}(m)}dm\,dk \\
&= 2\int \overline{\hat{v}(k)}\hat{v}(k) + \varepsilon \int \overline{\hat{v}(k)}\hat{b}(k)\hat{a}(k-m)\hat{v}(m)dm\,dk \\
&\quad +\varepsilon \int \hat{v}(m)\overline{\hat{b}(m)}\;\overline{\hat{a}(m-k)}\;\overline{\hat{v}(k)}dk\,dm \\
&= 2\int \overline{\hat{v}(k)}\hat{v}(k) + \varepsilon \int \overline{\hat{v}(k)}\hat{a}(k-m)\hat{v}(m)(\hat{b}(k) + \overline{\hat{b}(m)})dk\,dm
\end{aligned}$$

where we used $\hat{a}(\ell) = \overline{\hat{a}(-\ell)}$ due to the fact that a is real-valued.

Hence

$$2\|\hat{v}\|_{L^2}^2 \leq 2\|\hat{v}\|_{L^2}\|\hat{u}\|_{L^2} + \varepsilon s_1 ,$$

where with the Gagliardo-Nirenberg inequality

$$\begin{aligned}
s_1 &= \left|\int \overline{\hat{v}(k)}\hat{v}(m)\hat{a}(k-m)(\hat{b}(k) + \overline{\hat{b}(m)})dk\,dm\right| \\
&\leq \int |\overline{\hat{v}(k)}\hat{v}(m)||\hat{a}(k-m)|C|k-m|dm\,dk \\
&\leq \|v\|_{L^2}^2 \int |\hat{a}(\ell-m)|C|(\ell-m)|d\ell
\end{aligned}$$

since $|\hat{b}(k)+\overline{\hat{b}(m)}| = |\hat{b}(k)-\hat{b}(m)| \leq C|k-m|$ if $\hat{b}$ is Lipschitz-continuous and purely imaginary, and this last integral will be finite if the kernel a is sufficiently smooth.

If instead of inverting the transformation in the Sobolev spaces H^s, we work in the exponentially weighted spaces, $Y_{\sigma,s}$, the observation that $e^{\sigma|k|} \leq e^{\sigma|k-m|}e^{\sigma|m|}$, plus an argument very similar to that just above yields the estimates

$$\|u\|_{Y_{\sigma,s}} \leq C\|v\|_{Y_{\sigma,s+1}} \quad \text{and} \quad \|v\|_{Y_{\sigma,s}} \leq C\|u\|_{Y_{\sigma,s}} .$$

Now, we consider the first normal-form transformation constructed in the previous section:

$$\begin{aligned}
\tilde{R}_j^0 &= R_j^0 + \varepsilon N_j^{0,1}(\Psi, R^1) \\
\tilde{R}_j^1 &= R_j^1 + \varepsilon N_j^{1,1}(\Psi, R^1) + \varepsilon N_j^{1,0}(\Psi, R^0)
\end{aligned}$$

for $j = 1, 2, 3$. Recall that only the terms $N_j^{1,1}$ lose smoothness. Both $N_j^{0,1}$ and $N_j^{1,0}$ are bounded transformations from $Y_{\sigma,s}$ to $Y_{\sigma,s}$. Thus, we first consider just

$$\tilde{R}_j^1 = R_j^1 + \varepsilon N_j^{1,1}(\Psi, R^1) + \varepsilon N_j^{1,0}(\Psi, R^0) . \tag{89}$$

From the previous section we know that

$$\hat{N}_j^{1,1}(\Psi, R^1)(k) = \sum_{\substack{l=-1,1 \\ n=1,2,3}} \int \hat{n}_{j,l,n}^{1,1}(k)\hat{\psi}_l(k-m)\hat{R}_n^1(m)dm ,$$

where from the explicit formula (63) one can verify that $\hat{n}^{1,1}_{j,l,j}$ satisfies the hypotheses on the kernel assumed in the above model problem whereas the components $\hat{n}^{1,1}_{j,l,n}$ for $j \neq n$ do not cause a loss of regularity. Furthermore, $\hat{\psi}_\ell$ plays the role of $\hat{a}$ in the model problem and hence has the necessary smoothness properties. Thus, we multiply both sides of (89) by R^1_j, add together the cases for $j = 1, 2, 3$ and take the $Y^R_{\sigma,s}$ norm of both sides. Then we find analogously to the above model problem that

$$\|R^1\|^2_{Y^R_{\sigma,s}} \leq \|R^1\|_{Y^R_{\sigma,s}} \|\tilde{R}^1\|_{Y^R_{\sigma,s}} + C_1\varepsilon \|R^1\|^2_{Y^R_{\sigma,s}} + C_2\varepsilon(\|R^1\|^2_{Y^R_{\sigma,s}} + \|R^0\|^2_{Y^R_{\sigma,s}}) \ , \tag{90}$$

where $\|R^1\|^2_{Y^R_{\sigma,s}} = \|R^1_1\|^2_{Y^R_{\sigma,s}} + \|R^1_2\|^2_{Y^R_{\sigma,s}}$, and similarly for $\|\tilde{R}^1\|_{Y^R_{\sigma,s}}$ and $\|R^0\|_{Y^R_{\sigma,s}}$.

This inequality implies that the transformation $R^1 \to \tilde{R}^1$ is 1–1, hence invertible and satisfies the estimate

$$\|R^1\|^2_{Y^R_{\sigma,s}} \leq \left(\frac{1}{1 - C_3\varepsilon}\right) \left(\|\tilde{R}^1\|^2_{Y^R_{\sigma,s}} + \varepsilon C_4 \|R^0\|^2_{Y^R_{\sigma,s}}\right) \tag{91}$$

so that we can write

$$R^1_j = \tilde{R}^1_j + \varepsilon F(\tilde{R}^1, R^0). \tag{92}$$

We now consider the transformation for R^0_j, which with the help of (91). We can write

$$\begin{aligned} \widetilde{R}^0_j &= R^0_j + \varepsilon N^{0,1}_j(\Psi, R^1) \\ &= R^0_j + \varepsilon N^{0,1}_j(\Psi, \tilde{R}^1) + \varepsilon^2 N^{0,1}_j(\Psi, F(\tilde{R}^1, R^0)), \end{aligned} \tag{93}$$

or

$$R^0_j = (\tilde{R}^0_j - \varepsilon N^{0,1}_j(\Psi, \tilde{R}^1)) - \varepsilon^2 N^{0,1}_j(\Psi, F(\tilde{R}^1, R^0)) \ . \tag{94}$$

Recall that $N^{0,1}$ is smoothing as we remarked in (57) and the extra power of ε insures that $\varepsilon^2 N^{0,1}(\Psi, F(\tilde{R}^1, R^0))$ is also small. Thus (94) can be inverted by a Neumann series and we see that the normal-form transformation (52) is invertible and satisfies the estimates claimed in Proposition 3.10.

5 The error estimates

In this final section, we verify that the difference between the true solution of the water wave problem and the (improved) NLS approximation remains small over the relevant time scale.

In order to solve and control the error equation (88), we use energy estimates in a scale of Banach spaces of analytic functions. Because we cut-off the Fourier transforms of our approximation functions in Fourier space (see (37)) the approximation functions and the residual term computed from them are analytic in a strip of width $\mathcal{O}(1)$ in the complex plane, even though our original solutions of the NLS equation were only in H^{s_A}.

We now use this analyticity to allow us to apply results on optimal regularity for *parabolic* equations. We do this by allowing the width of the domain of analyticity to shrink with time. This adds an "artificial" smoothing to the equation (88).

To see how rapidly we can allow the width of the analyticity strip to shrink note that it is initially of width $2\sigma = \mathcal{O}(1)$ and we need to control solutions of the error equation for times of $\mathcal{O}(\varepsilon^{-2})$ so we can shrink the width of the analyticity strip with a velocity of order $\mathcal{O}(\varepsilon^2)$.

Hence, we define

$$\widehat{\mathcal{R}}(k,t) = \widehat{S}(k,t)\widehat{w}(k,t) = \widehat{w}(k,t)e^{-|k|(a-b\varepsilon^2 t)}$$

with constants $a, b > 0$ chosen below. If $w(t) \in L^2$, then $\mathcal{R}(t)$ is analytic in a strip of width $a - b\varepsilon^2 t$, i.e., $t \in [0, a/(b\varepsilon^2)]$. Computing the equation for w we find

$$\partial_t w = \Lambda w - |k| b\varepsilon^2 w + \varepsilon^2 \widetilde{\ell}(w) + \varepsilon^3 \widetilde{G}(w) + \varepsilon^{-3}\vartheta^{-1}\widetilde{\mathrm{Res}}(\varepsilon\Psi), \tag{95}$$

where $\widetilde{\ell}(w) = S^{-1}(t)\ell(S(t)w)$, $\widetilde{G}(w) = S^{-1}(t)G(S(t)w)$, and $\widetilde{\mathrm{Res}}(\varepsilon\Psi) = S^{-1}(t)\mathrm{Res}(\varepsilon\Psi)$.

If we use the estimates on ℓ and G from Proposition 3.16, along with the fact that the support of $\mathrm{Res}(\varepsilon\Psi)$ is bounded in Fourier space, then we immediately obtain the following estimates for the terms in (95).

Corollary 5.1. *For any $r \geq 3$, there exist constants $\widetilde{C}_L$, $\widetilde{C}_G$ and $\widetilde{C}_R$ such that*

$$\begin{aligned} \|\widetilde{\ell}(w)\|_{H_R^{r-1}} &\leq \widetilde{C}_L \|w\|_{H_R^r}, \\ \|\widetilde{G}(w)\|_{H_R^{r-1}} &\leq \widetilde{C}_G \|w\|_{H_R^r} \|w\|_{H_R^{r-1}}, \\ \|\varepsilon^{-3}\vartheta^{-1}\widetilde{\mathrm{Res}}(\varepsilon\Psi)\|_{H_R^r} &\leq \widetilde{C}_R \varepsilon^2, \end{aligned}$$

where $H_R^r = H^r \times H^{r-1/2} \times H^{r-1/2}$.

We control the solutions of equation (95) using energy estimates and Gronwall's inequality. Fix some index $s \geq 6$ and define

$$\|f\|_{H^s}^2 = \|f\|_{L^2}^2 + \|f\|_{\mathring{H}^s}^2 \tag{96}$$

where

$$\|f\|_{\mathring{H}^s}^2 = \int |k|^{2s} |\hat{f}(k)|^2 dk. \tag{97}$$

We have

$$\begin{aligned} \frac{1}{2}\partial_t \|w_j\|_{L^2}^2 = -b\varepsilon^2 \int |k||\widehat{w}_j(k)|^2 dk + \varepsilon^2 \int |\widehat{w}_j(k)||(\widehat{\widetilde{\ell}(w)})_j(k)| dk \\ +\varepsilon^\beta \int |\widehat{w}_j(k)||(\widehat{\widetilde{G}(w)})_j(k)| dk + \int |\widehat{w}_j(k)||\varepsilon^{-3}\widehat{\vartheta}^{-1}(k)(\widehat{\widetilde{\mathrm{Res}}(w)})_j(k)| dk \end{aligned} \tag{98}$$

for $j = 1, 2, 3$. Applying the Cauchy-Schwarz inequality and the estimates of Corollary 5.1, we find

$$\begin{aligned} \frac{1}{2}\partial_t \|w_j\|_{L^2}^2 &\leq -b\varepsilon^2 \|w_j\|_{\mathring{H}^{1/2}}^2 + \|w_j\|_{L^2}(\widetilde{C}_L \varepsilon^2 \|w\|_{H_R^3} + \widetilde{C}_G \varepsilon^3 \|w\|_{H_R^2}\|w\|_{H_R^3} + \widetilde{C}_R \varepsilon^2) \\ &\leq -b\varepsilon^2 \|w_j\|_{\mathring{H}^{1/2}}^2 + \varepsilon^2(\widetilde{C}_L + \widetilde{C}_R)\|w\|_{H_R^3}^2 + \widetilde{C}_G \varepsilon^3 \|w\|_{H_R^3}^3 + \widetilde{C}_R \varepsilon^2. \end{aligned}$$

Now consider

$$\begin{aligned} \frac{1}{2}\partial_t \|w_j\|_{\mathring{H}^s}^2 = -b\varepsilon^2 \int |k|^{2s+1} |\widehat{w}_j(k)|^2 dk + \varepsilon^2 \int |k|^{2s} |\widehat{w}_j(k)||(\widehat{\widetilde{\ell}(w)})_j(k)| dk \\ +\varepsilon^\beta \int |k|^{2s}|\widehat{w}_j(k)||(\widehat{\widetilde{G}(w)})_j(k)| dk + \int |k|^{2s}|\widehat{w}_j(k)||\varepsilon^{-3}\widehat{\vartheta}^{-1}(k)(\widehat{\widetilde{\mathrm{Res}}(w)})_j(k)| dk. \end{aligned} \tag{99}$$

If we once again apply the Cauchy-Schwarz inequality and the estimates in Corollary 5.1 we can bound the last three integrals in (99) by

$$\|w_j\|_{\mathring{H}^{s+1/2}} \left\{ \widetilde{C}_L \varepsilon^2 \|w\|_{H_R^{s+1/2}} + \widetilde{C}_G \varepsilon^3 \|w\|_{H_R^{s-1/2}} \|w\|_{H_R^{s+1/2}} + \widetilde{C}_R \varepsilon^2 \right\}. \tag{100}$$

Combining (99) and (100) gives

$$\frac{1}{2}\partial_t\|w_j\|^2_{\mathring{H}^s} \leq -\varepsilon^2(b-(\widetilde{C}_L+\widetilde{C}_R)-\widetilde{C}_G\varepsilon\|w\|_{H_R^{s-1/2}})\|w_j\|^2_{H^{s+1/2}}+\widetilde{C}_R\varepsilon^2. \tag{101}$$

Combining this with the estimate on the L^2-norm of w and using $\|f\|_{H^3} \leq 2\|f\|_{H^r}$ for all $r \geq 3$ we obtain the inequality

$$\frac{1}{2}\partial_t\|w\|^2_{H_R^s} \leq -\varepsilon^2(b-3(\widetilde{C}_L+\widetilde{C}_R)-3\widetilde{C}_G\varepsilon\|w\|_{H_R^{s-1/2}})\|w\|^2_{H_R^{s+1/2}}+2\widetilde{C}_R\varepsilon^2. \tag{102}$$

Applying Gronwall's inequality to (102) we obtain:

Proposition 5.2. *If* $b-3(\widetilde{C}_L+\widetilde{C}_R)-3\widetilde{C}_G\varepsilon\sup_{0\leq t\leq t_0}\|w(t)\|_{H_R^{s-1/2}} \geq 0$, *then*

$$\sup_{0\leq t\leq t_0}\|w(t)\|^2_{H_R^s} \leq (\|w(0)\|^2_{H_R^s}+2\widetilde{C}_R\varepsilon^2 t_0).$$

Take $t_0=\varepsilon^{-2}\widetilde{T}_0$ and $\|w(0)\|^2_{H_R^s} \leq 2\widetilde{C}_R\widetilde{T}_0$. Then choose b such that $b-3(\widetilde{C}_L+\widetilde{C}_R)-24\widetilde{C}_G\widetilde{C}_R\widetilde{T}_0\varepsilon \geq 0$. The Proposition 5.2 implies

Corollary 5.3. *For all* $0 \leq \varepsilon^2 t \leq \widetilde{T}_0$,

$$\|w(t)\|^2_{H_R^s} \leq 4\widetilde{C}_R\widetilde{T}_0.$$

Finally we must check that the smoothing operator $S(t)$ is well defined. We require that the constants a and b in its definition be such that $\sigma > a$ and $a-b\varepsilon^2 t > a/2$ for all $0 \leq \varepsilon^2 t \leq \widehat{T}_0$. In this case $S(t)$ is well defined. (Note that this means in particular that $\widehat{T}_0 < \sigma/(2b)$.) Finally, we have

Corollary 5.4. *Choose* $T_1=\min(\widetilde{T}_0,\widehat{T}_0)$. *Then*

$$\begin{aligned}\sup_{0\leq\varepsilon^2 t\leq T_1}\|\mathcal{R}(t)\|_{Y^R_{a/2,s}} &\leq \sup_{0\leq\varepsilon^2 t\leq T_1}\|\mathcal{R}(t)\|_{Y^R_{a-b\varepsilon^2 t,s}} \leq \sup_{0\leq\varepsilon^2 t\leq\widetilde{T}_0}\|\mathcal{R}(t)\|_{Y^R_{a-b\varepsilon^2 t,s}} \\ &= \sup_{0\leq\varepsilon^2 t\leq\widetilde{T}_0}\|S(t)w(t)\|_{Y^R_{a-b\varepsilon^2 t,s}} = \sup_{0\leq\varepsilon^2 t\leq\widetilde{T}_0}\|w(t)\|_{H_R^s} \leq 4\widetilde{C}_R\widetilde{T}_0.\end{aligned} \tag{103}$$

Since the $Y^2_{a/2,s}$-norm controls any Sobolev norm, we obtain

Corollary 5.5. *Choose* $T_1=\min(\widetilde{T}_0,\widehat{T}_0)$. *Then*

$$\sup_{0\leq\varepsilon^2 t\leq T_1}\|\mathcal{R}(t)\|^2_{H_R^s} \leq 4\widetilde{C}_R\widetilde{T}_0.$$

Combining this estimate with Proposition 3.10, Proposition 3.15, and Lemma 2.6 implies Theorem 2.8.

A Appendix

A.1 Some estimates on the operator $\mathcal{K}(X)$

In this appendix we prove the statements about the analyticity of the operater $\mathcal{K}$ used in previous sections. Note that as described in Section 2 the value of $\mathcal{K}(X_1, X_2)U_1$ is obtained by solving the boundary value problem

$$\Delta\phi = 0, \quad \text{in} \quad \Omega(t), \tag{104}$$

$$\partial_{x_2}\phi = 0, \quad \text{for} \quad x_2 = -1, \tag{105}$$

$$\partial_{x_1}\phi = U_1, \quad \text{on} \quad \Gamma(t), \tag{106}$$

where $\Omega(t)$ is the domain $\{(x_1, x_2) \mid -\infty < x_1 < \infty \ , -1 < x_2 < \eta(x_1, t)\}$ and $\Gamma(t) = \{(x_1, \eta(x_1, t))| \ x_1 \in \mathbb{R}\}$ is the upper surface of the fluid, specified in Lagrangian variables by the curve $(\alpha + X_1(\alpha, t), X_2(\alpha, t))$. If $\phi(x_1, x_2)$ is the solution of this problem (for fixed t) then $\partial_{x_2}\phi|_{\Gamma(t)} = \mathcal{K}(X_1, X_2)U_1$.

We now solve this boundary value problem in the spaces $Y_{\sigma,s}$ to analyze the analyticity of $\mathcal{K}(X)$. We will reduce the problem on $\Omega(t)$ to a problem on the fixed rectangular domain $R = \{(x, y) \mid -\infty < x < \infty \ , 0 < y < 1\}$ and with this in mind we introduce the Banach spaces

$$\mathbb{K}^r_{\sigma,s} = H^r((0, 1), Y_{\sigma,s}) \ .$$

Before treating the full problem that defines $\mathcal{K}(X)$ we derive a pair of simple lemmas that we will use later.

Let $u(x, y) = \partial_x\phi(x, y)$ and $v(x, y) = \partial_y\phi(x, y)$. Consider the homogeneous boundary value problem

$$\left.\begin{array}{rcl} \partial_x u + \partial_y v &=& 0 \ , \\ \partial_y u - \partial_x v &=& 0 \ , \end{array}\right\} \qquad (x, y) \in R \ ,$$

$$v|_{y=0} = 0 \ , \qquad u(x, 1) = U(x) \ . \tag{107}$$

Lemma A.1. *If $U \in Y_{\sigma,s}$ then $u \in \mathbb{K}^0_{\sigma,s+(1/2)} \cap \mathbb{K}^1_{\sigma,s-(1/2)}$.*

Proof. Taking Fourier transforms with respect to x we find

$$\hat{u}(k, y) = \frac{\hat{U}(k)\cosh(ky)}{\cosh(k)} \ .$$

Then

$$\|u\|^2_{\mathbb{K}^0_{\sigma,s+(1/2)}} = \int_{k=-\infty}^{\infty}\int_{y=0}^{1} e^{2\sigma|k|}(1 + k^2)^{s+(1/2)} \left|\frac{\hat{U}(k)\cosh(ky)}{\cosh(k)}\right|^2 dydk \ .$$

Performing the integral with respect to y the integrand becomes

$$e^{2\sigma|k|}(1 + k^2)^{s+(1/2)}|\hat{U}(k)|^2 \left|\frac{\cosh(k)\sinh(k) + k}{2k\cosh(k)}\right|^2 \leq Ce^{2\sigma|k|}\frac{(1 + k^2)^{s+(1/2)}}{(1 + |k|)^2}|\hat{U}(k)|^2$$

and the integral over k is finite since $U \in Y_{\sigma,s}$. A similar calculation shows $u \in \mathbb{K}^1_{\sigma,s-(1/2)}$. □

Next consider the inhomogeneous system of equations

$$\partial_x u + \partial_y v \ \ = \ \ f \ ,$$

$$\begin{array}{rcl} \partial_y u - \partial_x v &=& g \ , \\ v|_{y=0} = u|_{y=1} &=& 0 \ . \end{array} \tag{108}$$

If we represent $u(x,y)$ as a series in $\cos((2m+1)\pi y/2)$ and $v(x,y)$ as a series in $\sin((2m+1)\pi x/2)$ then a computation similar to that above yields:

Lemma A.2. *Suppose f and g are elements of $\mathbb{K}^0_{\sigma,s} \cap \mathbb{K}^1_{\sigma,s-1}$. Then the solutions u and v of the inhomogeneous system are elements of $\mathbb{K}^0_{\sigma,s+1} \cap \mathbb{K}^1_{\sigma,s}$.*

With these preliminaries in hand we now turn to a consideration of the operator $\mathcal{K}(X)$. We map the variables (x_1, x_2) in the original fluid domain onto a rectangle via the change of variables

$$x_1 = \alpha + X_1(\alpha) \ , \quad x_2 = z(1 + X_2(\alpha)) \ .$$

If $u(x_1, x_2) = \tilde{u}(\alpha, z)$ and $v(x_1, x_2) = \tilde{v}(\alpha, z)$ then

$$\begin{array}{rcl} \partial_\alpha \tilde{u} + \partial_z \tilde{v} &=& R_1 \ , \\ \partial_z \tilde{u} - \partial_\alpha \tilde{v} &=& R_2 \ , \end{array} \tag{109}$$

where R_1 and R_2 are given by

$$R_1 = -\left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z \tilde{v} + \left(\frac{X_2}{1+X_2}\right)\partial_z \tilde{v} + \left(\frac{z\partial_\alpha X_2 \partial_z \tilde{u}}{(1+X_2)}\right) \ ,$$

and

$$R_2 = \left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z \tilde{u} + \left(\frac{X_2}{1+X_2}\right)\partial_z \tilde{u} - \left(\frac{z\partial_\alpha X_2 \partial_z \tilde{v}}{(1+X_2)}\right) \ ,$$

subject to the boundary conditions $\tilde{u}(\alpha, 1) = \partial_t X_1$ and $\tilde{v} = 0$.

Let (u^h, v^h) be the solution of the homogeneous equations:

$$\begin{array}{rcl} \partial_\alpha u^h + \partial_z v^h &=& 0 \ , \\ \partial_z u^h - \partial_\alpha v^h &=& 0 \end{array} \tag{110}$$

with boundary conditions $u^h(\alpha, 1) = U_1$ and $v^h = 0$.

We can solve this problem with the aid of Lemma A.1 and we find:

Lemma A.3. *If $U_1 \in Y_{\sigma,s-(1/2)}$ the u^h and v^h are in $\mathbb{K}^0_{\sigma,s} \cap \mathbb{K}^1_{\sigma,s-1}$.*

Remark A.4. *Note that the boundary value of $v^h|_{z=1}$ gives us the value of the linearized operator $\mathcal{K}_0 U_1$. Applying the trace theorem we see that Lemma A.3 implies that $\mathcal{K}_0$ is a bounded operator from $Y_{\sigma,s-(1/2)}$ to itself.*

Now set $\overline{u} = \tilde{u} - u^h$, $\overline{v} = v - v^h$ and we find that

$$\begin{array}{rl} \partial_\alpha \bar{u} + \partial_z \bar{v} &= \bar{R}_1 \ , \\ \partial_z \bar{u} - \partial_\alpha \bar{v} &= \bar{R}_2 \ , \end{array} \tag{111}$$

where R_1 and R_2 are given by

$$\begin{array}{rcl} \bar{R}_1 &=& -\left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z \bar{v} + \left(\frac{X_2}{1+X_2}\right)\partial_z \bar{v} + \left(\frac{z\partial_\alpha X_2 \partial_z \bar{u}}{(1+X_2)}\right) + \partial_\alpha u^h + \partial_z v^h \\ && -\left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z v^h + \left(\frac{X_2}{1+X_2}\right)\partial_z v^h + \left(\frac{z\partial_\alpha X_2 \partial_z u^h}{(1+X_2)}\right) \ , \end{array} \tag{112}$$

and

$$\bar{R}_2 = \left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z\bar{u} + \left(\frac{X_2}{1+X_2}\right)\partial_z\bar{u} - \left(\frac{z\partial_\alpha X_2\partial_z\bar{v}}{(1+X_2)}\right) + \partial_z u^h - \partial_\alpha v^h + \left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z u^h + \left(\frac{X_2}{1+X_2}\right)\partial_z u^h - \left(\frac{z\partial_\alpha X_2\partial_z v^h}{(1+X_2)}\right)\,, \tag{113}$$

where now we have zero boundary conditions $\overline{u}(\alpha,1)=0$ and $\overline{v}(\alpha,0)=0$.

Define

$$\mathbf{F}(\bar{u},\bar{v};\partial_\alpha X_1, X_2) = (F_1(\bar{u},\bar{v};\partial_\alpha X_1,X_2), F_2(\bar{u},\bar{v};\partial_\alpha X_1,X_2))\,,$$

where $F_1 = \partial_\alpha\bar{u} + \partial_z\bar{v} - \bar{R}_1$ and $F_2 = \partial_z\bar{u} - \partial_\alpha\bar{v} = \bar{R}_2$.

Note that a solution $(\bar{u},\bar{v})$ of our partial differential equations is a zero of $\mathbf{F}$. We will solve the PDE's by applying the implicit function theorem to find zeros of $\mathbf{F}$.

Define the Banach spaces $E = (\mathbb{K}^0_{\sigma,s}\cap\mathbb{K}^1_{\sigma,s-1})^2$ with boundary conditions $\bar{u}(\alpha,1)=0$ and $\bar{v}(\alpha,0)=0$, $F = Y_{\sigma,s-1}\times Y_{\sigma,s}$ and $G = (\mathbb{K}^0_{\sigma,s-1})^2$. (In fact we consider the complex extensions of these spaces so that we can work with complex Banach spaces.)

Note that $\mathbf{F}$ is an analytic function from $E\times F$ into G.

If we take $X_1 = X_2 = 0$, then we can find (u^0,v^0) such that

$$\mathbf{F}(u^0,v^0;0,0) = (0,0)$$

since this is just the solution of the inhomogeneous, partial differential equations. More precisely, this can be rewritten as

$$\begin{aligned}
\partial_\alpha u^0 + \partial_z v^0 &= \partial_z u^h - \partial_\alpha v^h + \left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z u^h + \left(\frac{X_2}{1+X_2}\right)\partial_z u^h \\
&\quad - \left(\frac{z\partial_\alpha X_2\partial_z v^h}{(1+X_2)}\right)\,, \\
\partial_z u^0 - \partial_\alpha v^0 &= \partial_z u^h - \partial_\alpha v^h + \left(\frac{\partial_\alpha X_1}{1+X_2}\right)\partial_z u^h + \left(\frac{X_2}{1+X_2}\right)\partial_z u^h \\
&\quad - \left(\frac{z\partial_\alpha X_2\partial_z v^h}{(1+X_2)}\right)\,.
\end{aligned} \tag{114}$$

Note that the right hand side of this system of equations is an element of the Banach space G and using Fourier transform we can solve for $(u^0,v^0)\in E$.

Next observe that if we linearize $\mathbf{F}$ at $(u^0,v^0;0,0)$ we have

$$\left(D_{(u,v)}\mathbf{F}(u^0,v^0;0,0)\right)\begin{pmatrix}U\\V\end{pmatrix} = \begin{pmatrix}\partial_\alpha U + \partial_z V\\ \partial_z U - \partial_\alpha V\end{pmatrix}\,.$$

But then, for any $(f,g)\in G$, we see that

$$\left(D_{(u,v)}\mathbf{F}(u^0,v^0;0,0)\right)\begin{pmatrix}U\\V\end{pmatrix} = \begin{pmatrix}\partial_\alpha U + \partial_z V\\ \partial_z U - \partial_\alpha V\end{pmatrix} = \begin{pmatrix}f\\g\end{pmatrix}$$

again just requires us to solve a the same linear, constant coefficient partial differential equation. Again, for any $(f,g)\in G$ we can find $(U,V)\in E$ by Fourier transform and hence $D_{(u,v)}\mathbf{F}(u^0,v^0;0,0)$ has a bounded inverse.

Thus, we can apply the analytic implicit function theorem and we find that for any $(\partial_\alpha X_1, X_2)$ in a sufficiently small neighborhood of the origin in F there exist solutions $(\bar{u}, \bar{v}) \in E$ depend analytically on $\partial_\alpha X_1$, X_2.

Since the trace operator is a bounded linear operator from E into $Y_{\sigma,s-(1/2)}$ we see that $v(\alpha, 1) = \mathcal{K}(X_1, X_2)(\partial_t X_1)$ depends analytically on $(\partial_\alpha X_1, X_2)$ in a sufficiently small neighborhood of the origin in F.

Thus, we have proven:

Proposition A.5. *$\mathcal{K}(X_1, X_2)$ is a linear operator from $Y_{\sigma,s-(1/2)}$ to itself which depends analytically on $(\partial_\alpha X_1, X_2) \in Y_{\sigma,s} \times Y_{\sigma,s}$.*

One other operator which we used in Section 2 was $\mathcal{M}_1$. Recall that to avoid the secular growth in the variable X_1, we introduced the variable $Z_1 = \mathcal{K}_0 X_1$ and we associated to Z_1 the operator

$$\mathcal{M}_1(Z_1, \cdot) = [X_1, \mathcal{K}_0]\cdot$$

which satisfies

Lemma A.6. *Let $r \geq 0$, $q > 1/2$ and $0 \leq p \leq q$. Then there exists a $C > 0$ such that*

$$\|\mathcal{M}_1(a, u)\|_{H^r} \leq C\|a\|_{H^{r+p}}\|u\|_{H^{q-p}},$$

$$\|\mathcal{M}_1(a, u)\|_{H^r} \leq C\|\hat{a}\|_{L^1(r+p)}\|u\|_{H^{q-p}}.$$

Proof. See [SW00, Corollary 3.13] and [SW03, Remark A.6] . □

Remark A.7. *$\mathcal{M}_1$ is well defined, even though $\mathcal{K}_0$ is not invertible in general, due to the commutator in its definition.*

In order to express the term $\partial_\alpha X_1$ in terms of Z_1 we defined additionally the operator

$$\mathcal{M}_2 \cdot = -\partial_\alpha (\mathcal{K}_0)^{-1} \cdot$$

which is a map from H^{s+1} to H^s.

Remark A.8. *Finally, the operator $(1 + \mathcal{K}_0^2)\cdot$ is infinitely smoothing due to the fact that in Fourier space its symbol $(1 + \hat{\mathcal{K}}_0(k)^2)$ vanishes with some exponential rate for $|k| \to \infty$.*

A.2 Some properties of our function spaces.

It is more or less obvious that the spaces $Y_{\sigma,s}$ are Banach spaces. For $s > 1/2$ something stronger is true.

Lemma A.9. *The spaces $Y_{\sigma,s}$ are Banach algebras for all $\sigma \geq 0$ and all $s > \frac{1}{2}$.*

Proof. Suppose that u and v are in $Y_{\sigma,s}$. Then

$$\begin{aligned}
\|uv\|_{\sigma,s}^2 &= \int (1+k^2)^s e^{2\sigma|k|} \left(\int \hat{u}(k-\ell)\hat{v}(\ell) d\ell \right)^2 dk \\
&\leq C \int \left(\int [(1+|k-\ell|^2)^{s/2} + (1+\ell^2)^{s/2}] e^{\sigma(|k|+|k-\ell|)} |\hat{u}(k-\ell)||\hat{v}(\ell)| d\ell \right)^2 dk \\
&\leq C \int \left(\int (1+|k-\ell|^2)^{s/2} e^{\sigma(|k|+|k-\ell|)} |\hat{u}(k-\ell)||\hat{v}(\ell)| d\ell \right)^2 dk \\
&\quad + C \int \left(\int (1+|k|^2)^{s/2} e^{\sigma(|k|+|k-\ell|)} |\hat{u}(k-\ell)||\hat{v}(\ell)| d\ell \right)^2 dk \qquad (115)
\end{aligned}$$

Each of the two terms in the last line of this inequality can be interpreted as the square of the L^2-norm of a convolution and hence we use Young's inequality to bound each of them in turn. For instance the first is bounded by

$$\begin{aligned} &C\left(\int(1+|\ell|^2)^s e^{2\sigma|\ell|}|\hat{v}(\ell)|^2 d\ell\right)\left(\int e^{2\sigma|k|}|\hat{u}(k)dk\right) \\ &\quad = C\|v\|_{\sigma,s}\left(\int(1+|k|^2)^{-s/2}(1+|k|^2)^{s/2}e^{2\sigma|k|}|\hat{u}(k)|dk\right) \end{aligned} \tag{116}$$

If we now apply the Cauchy-Schwarz inequality to this last integral we find that it is bounded by a constant times $\|u\|_{\sigma,s}$ provided $s > 1/2$. Applying a similar argument to the second term in the last expression in (115) completes the proof of the lemma. □

A.3 Explicit form of the bilinear terms in our equations.

In this appendix we give explicit formulas for the bilinear terms appearing in the equations of motion (and which are important for analyzing the normal-form transformation):

$$B_1(c_1, c_2) = -\mathcal{M}_1(c_1, \partial_\alpha c_2) \ , \tag{117}$$

$$B_2(c_1, c_3) = -\mathcal{M}_1(c_1, \partial_\alpha c_3) \ , \tag{118}$$

$$\begin{aligned} B_3(c_2, c_3) \;&=\; \mathcal{M}_1(sc_3, \partial_\alpha c_2) - \mathcal{M}_1(sc_2, \partial_\alpha c_3) \\ &\quad -[sc_3 + \mathcal{K}_0(sc_3)\mathcal{K}_0]\partial_\alpha c_2 + [sc_2 + \mathcal{K}_0(sc_2)\mathcal{K}_0]\partial_\alpha c_3 \ , \end{aligned} \tag{119}$$

$$B_4(f, g) = -\mathcal{M}_1(sf, \partial_\alpha g) + [sf + \mathcal{K}_0(sf)\mathcal{K}_0]\partial_\alpha g \ , \tag{120}$$

$$B_5(c_3, c_3) = \mathcal{M}_1(sc_3, \partial_\alpha c_3) - [sc_3 + \mathcal{K}_0(sc_3)\mathcal{K}_0]\partial_\alpha c_3 \ , \tag{121}$$

$$B_6(c_1, c_2) = \frac{1}{2s}\mathcal{M}_1(c_1, \partial_\alpha c_2) - \frac{1}{2}\mathcal{M}_2(c_1)\partial_\alpha(sc_2) \ , \tag{122}$$

$$B_7(c_1, c_3) = \frac{1}{2s}\mathcal{M}_1(c_1, \partial_\alpha c_3) + \frac{1}{2}\mathcal{M}_2(c_1)\partial_\alpha(sc_3) \ , \tag{123}$$

$$\begin{aligned} B_8(c_2, c_3) \;&=\; \frac{1}{2s}\mathcal{M}_1(sc_2, \partial_\alpha c_3) - \frac{1}{2s}\mathcal{M}_1(sc_3, \partial_\alpha c_2) \\ &\quad -[sc_2 + \mathcal{K}_0(sc_2)\mathcal{K}_0]\partial_\alpha c_3 + [sc_3 + \mathcal{K}_0(sc_3)\mathcal{K}_0]\partial_\alpha c_2 \\ &\quad +\frac{1}{2}\mathcal{M}_2(sc_2)\partial_\alpha(sc_3) + \frac{1}{2}\mathcal{M}_2(sc_3)\partial_\alpha(sc_2) \\ &\quad -\frac{1}{2}(\partial_\alpha(sc_2))\mathcal{K}_0\partial_\alpha(sc_3) - \frac{1}{2}(\partial_\alpha(sc_3))\mathcal{K}_0\partial_\alpha(sc_2) \ , \end{aligned} \tag{124}$$

$$\begin{aligned} B_9(f, g) \;&=\; \frac{1}{2s}\mathcal{M}_1(sf, \partial_\alpha g) - [sf + \mathcal{K}_0(sf)\mathcal{K}_0]\partial_\alpha g \\ &\quad -\frac{1}{2}\mathcal{M}_2(sf)\partial_\alpha(sg) + \frac{1}{2}(\partial_\alpha(sf))\mathcal{K}_0\partial_\alpha(sg) \ , \end{aligned} \tag{125}$$

$$
\begin{aligned}
B_{10}(c_3,c_3) &= -\frac{1}{2s}\mathcal{M}_1(sc_3,\partial_\alpha c_3)+[sc_3+\mathcal{K}_0(sc_3)\mathcal{K}_0]\partial_\alpha c_3 \\
&\quad -\frac{1}{2}\mathcal{M}_2(sc_3)\partial_\alpha(sc_3)+\frac{1}{2}(\partial_\alpha(sc_3))\mathcal{K}_0\partial_\alpha(sc_3)\ ,
\end{aligned} \tag{126}
$$

$$
B_{11}(c_1,c_2)=-\frac{1}{2s}\mathcal{M}_1(c_1,\partial_\alpha c_2)-\frac{1}{2}\mathcal{M}_2(c_1)\partial_\alpha(sc_2)\ , \tag{127}
$$

$$
B_{12}(c_1,c_3)=-\frac{1}{2s}\mathcal{M}_1(c_1,\partial_\alpha c_3)+\frac{1}{2}\mathcal{M}_2(c_1)\partial_\alpha(sc_3)\ , \tag{128}
$$

$$
\begin{aligned}
B_{13}(c_2,c_3) &= -\frac{1}{2s}\mathcal{M}_1(sc_2,\partial_\alpha c_3)+\frac{1}{2s}\mathcal{M}_1(sc_3,\partial_\alpha c_2) \\
&\quad -[sc_2+\mathcal{K}_0(sc_2)\mathcal{K}_0]\partial_\alpha c_3+[sc_3+\mathcal{K}_0(sc_3)\mathcal{K}_0]\partial_\alpha c_2 \\
&\quad +\frac{1}{2}\mathcal{M}_2(sc_2)\partial_\alpha(sc_3)+\frac{1}{2}\mathcal{M}_2(sc_3)\partial_\alpha(sc_2) \\
&\quad -\frac{1}{2}(\partial_\alpha(sc_2))\mathcal{K}_0\partial_\alpha(sc_3)-\frac{1}{2}(\partial_\alpha(sc_3))\mathcal{K}_0\partial_\alpha(sc_2)\ ,
\end{aligned} \tag{129}
$$

$$
\begin{aligned}
B_{14}(f,g) &= -\frac{1}{2s}\mathcal{M}_1(sf,\partial_\alpha g)-[sf+\mathcal{K}_0(sf)\mathcal{K}_0]\partial_\alpha g \\
&\quad -\frac{1}{2}\mathcal{M}_2(sf)\partial_\alpha(sg)+\frac{1}{2}(\partial_\alpha(sf))\mathcal{K}_0\partial_\alpha(sg)\ ,
\end{aligned} \tag{130}
$$

$$
\begin{aligned}
B_{15}(c_3,c_3) &= \frac{1}{2s}\mathcal{M}_1(sc_3,\partial_\alpha c_3)+[sc_3+\mathcal{K}_0(sc_3)\mathcal{K}_0]\partial_\alpha c_3 \\
&\quad -\frac{1}{2}\mathcal{M}_2(sc_3)\partial_\alpha(sc_3)+\frac{1}{2}(\partial_\alpha(sc_3))\mathcal{K}_0\partial_\alpha(sc_3)\ .
\end{aligned} \tag{131}
$$

References

[AS81] Ablowitz, M.J., Segur, H.: *Solitons and the inverse scattering transform.* SIAM Studies in Applied Mathematics, 4. Philadelphia: SIAM - Society for Industrial and Applied Mathematics. (1981)

[Am03] Ambrose, D.M.: Well-posedness of vortex sheets with surface tension. *SIAM J. Math. Anal.* **35**, 211-244 (2003)

[AM05] Ambrose, D.M., Masmoudi, N.: The zero surface tension limit of two–dimensional water waves. *Comm. Pure Appl. Math.* **58**, 1287-1315 (2005)

[AM09] Ambrose, D.M., Masmoudi, N.: The zero surface tension limit of three–dimensional water waves. *Indiana Univ. Math. J.* **58**(2), 479-521 (2009)

[Cr85] Craig, W.: An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits. *Comm. Partial Differential Equations.* **10**(8), 787-1003 (1985)

[CSS92] Craig, W., Sulem, C., Sulem, P.L.: Nonlinear modulation of gravity waves: a rigorous approach. *Nonlinearity* **5**, 497-552 (1992)

[CSS97] Craig, W., Schanz, U., Sulem, C.: The modulational limit of three-dimensional water waves and the Davey-Stewartson system. *Annales de l'IHP: Analyse Nonlinéaire* **14**, 615-667 (1997)

[DS06] Düll, W.-P., Schneider, G.: Justification of the Nonlinear Schrödinger equation for a resonant Boussinesq model. *Indiana Univ. Math. J.* **55**(6), 1813-1834 (2006)

[DS71] Dunford, N., Schwartz, J.T.: *Linear operators.* Part I-III. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988. Reprint of the 1958-1971 original.

[GS01] Gallay, T., Schneider, G.: KP description of unidirectional long waves. The model case. *Proc. Roy. Soc. Edinburgh Sect. A* **131**, 885-898 (2001)

[GMS12] Germain, P., Masmoudi, N., Shatah, J.: Global solutions for the gravity water waves equation in dimension 3. *Ann. of Math.* **175**(2), 691-754 (2012)

[Ig01] Iguchi, T.: Well-posedness of the initial value problem for capillary-gravity waves. *Funkcial. Ekvac.* **44**, 219-241 (2001)

[Kal88] Kalyakin, L.A.: Asymptotic decay of a one-dimensional wave packet in a nonlinear dispersive medium. *Math. USSR Sbornik Surveys* **60**(2), 457-483 (1988)

[KN79] Kano, T. Nishida, T.: Sur les ondes de surface de l'eau avec une justification mathematique des equations des ondes en eau profonde. *J. Math. Kyoto Univ.* **19**, 335-370 (1979)

[Kat75] Kato, T.: Quasi-linear equations of evolution with applications to partial differential equations. *LNM* **448**, 25-70, Springer 1974

[KSM92] Kirrmann, P., Schneider, G., Mielke, A.: The validity of modulation equations for extended systems with cubic nonlinearities. *Proc. Roy. Soc. Edinburgh Sect. A* **122**, 85-91 (1992)

[La05] Lannes, D.: Well-Posedness of the Water-Waves Equations. *J. Amer. Math. Soc.* **18**(3), 605-654 (2005)

[Na74] Nalimov, V.I.: The Cauchy-Poisson problem. (in russian), *Dynamika Splosh. Sredy* **18**, 104-210 (1974)

[OOS00] Osborne, A., Onorato, M., Serio, M.: The nonlinear dynamics of rogue waves and holes in deep-water gravity wave trains. *Phys. Lett. A* **275**(5-6), 386-393 (2000)

[Schn95] Schneider, G.: Validity and Limitation of the Newell-Whitehead equation. *Math. Nachr.* **176**, 249-263 (1995)

[Schn98a] Schneider, G.: Approximation of the Korteweg-de Vries equation by the Nonlinear Schrödinger equation. *J. Differential Equations* **147**, 333-354 (1998)

[Schn98b] Schneider, G.: Justification of modulation equations for hyperbolic systems via normal forms. *NoDEA Nonlinear Differential Equations Appl.* **5**, 69-82 (1998)

[Schn05] Schneider, G.: Justification and failure of the Nonlinear Schrödinger equation in case of non-trivial quadratic resonances. *J. Differential Equations* **216**(2), 354-386 (2005)

[SW00] Schneider, G., Wayne, C.E.: The long wave limit for the water wave problem. I. the case of zero surface tension. *Comm. Pure Appl. Math.* **53**(12), 1475-1535 (2000)

[SW02] Schneider, G., Wayne, C.E.: The rigorous approximation of long-wavelength capillary-gravity waves. *Arch. Rat. Mech. Anal.* **162**, 247-285 (2002)

[SW03] Schneider, G., Wayne, C.E.: Estimates for the three wave interaction of surface water waves. *European J. Appl. Math.* **14**, 547-570 (2003)

[SW11] Schneider, G., Wayne, C.E.: Justification of the NLS approximation for a quasi-linear water wave model. *J. Differential Equations* **251**(2), 238-269 (2011)

[Sh76] Shinbrot, M.: The initial value problem for surface waves under gravity, I. The simplest case. *Indiana Univ. Math. J.* **25**, 281-300 (1976)

[Wu97] Wu, S.: Well-posedness in Sobolev-spaces of the full water wave problem in 2D. *Invent. Math.* **130**, 39-72 (1997)

[Wu99] Wu, S.: Well-posedness in Sobolev-spaces of the full water wave problem in 3D. *J. Amer. Math. Soc.* **12**, 445-495 (1999)

[Wu09] Wu, S.: Almost global wellposedness of the 2-D full water wave problem. *Invent. Math.* **177**(1), 45-135 (2009)

[Wu11] Wu, S.: Global well-posedness of the 3-D full water wave problem. *Invent. Math.* **184**(1), 125-220 (2011)

[TW12] Totz, N., Wu, S.: A rigorous justification of the modulation approximation to the 2D full water wave problem. *Comm. Math. Phys.*, accepted (2012).

[Yo82] Yosihara, H.: Gravity waves on the free surface of an incompressible perfect fluid of finite depth. *RIMS Kyoto* **18**, 49-96 (1982)

[Yo83] Yosihara, H.: Capillary-gravity waves for an incompressible ideal fluid. *J. Math. Kyoto Univ.* **23**, 649-694 (1983)

[Za68] Zakharov, V.E.: Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Sov. Phys. J. Appl. Mech. Tech. Phys* **4**, 190-194 (1968)